CHARACTERIZATION OF A PLANT NUCLEAR PROTEIN THAT BINDS A SPECIFIC SEQUENCE OF THE *780* GENE PROMOTER OF T-DNA

BY

ELOISE CHRISTINE ADAMS

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1993

ACKNOWLEDGEMENTS

I wish to acknowledge all of my colleagues in the Gurley lab for their support and the friendship given to me during my tenure as a graduate student. Special thanks goes to Dulce Barros who gave of her time and her expertise in the beginning and at the end. To Kevin, my lab mate, I extend my thanks for the inspiration and stimulating conversation. To the mentor of all mentors, William B. Gurley, I say thanks for his support, guidance, and most of all his skepticism. I also owe a debt of gratitude to my committee members, Robert Ferl, Francis Davis, Don McCarty and Robert Schmidt, for their support and encouragement. To my husband, Howard and my daughter, Stephanie, I express a loving thanks for their support, inspiration and confidence. And I would like to acknowledge the McKnight Black Doctoral Fellowship for making this all possible.

TABLE OF CONTENTS

ABBREVIATIONS

Adh1	alcohol dehydrogenase gene
ags	agropine synthase gene
as-1	activation sequence
ASF-1	Activation Sequence Factor
ATcom	AT composite
AT-1	pea nuclear factor that binds AT-rich sequences
bp	base pair
b-Zip	basic leucine zipper
cab	chlorophyll a/b binding protein gene
CAF	CA binding factor
CaMV 35S	cauliflower mosaic virus 35S
CAT	chloramphenicol acetyltransferase
chs	chalcone synthase gene
CNBr	cyanogen bromide
COUP-TF	chicken ovalbumin upstream promoter transcription factor
CPRF-1, 2, 3	Common Plant Regulatory Factor
DMS	dimethyl sulfate
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EGTA	(β-amino ethyl ether)-N; N, N', N'-tetraacetic acid
EmBP-1	wheat Em binding protein (Em = embryo)
GATA-1	GATA binding protein
GBF	G-box binding factor
Gmhsp17.5E	soybean small molecular weight heat shock gene
GT-1	binds light responsive elememts box II and box III (GT-rich sequences)
GUS	β-glucuronidase
HBP-1	wheat histone binding protein
HEPES	N-(2-Hydroxyethyl)piperazine-N-(2-ethanesulfonic acid)
H4TF-1	histone 4 transcription factor
HSE	heat shock element
HSF	heat shock transcription factor
kD	kilodalton
LRE	light responsive element
Le1	soybean lectin gene
mas	mannopine synthase
NEM	N-ethylmaleimide
nos	nopaline synthase gene
NP-40	nonionic detergent 40
ocs	octopine synthase gene
OCSBF	*ocs*-element binding factor

OCSTF	ocs-element transcription factor
ons	octopine, nopaline secretion
pal	phenylalanine ammonia-lyase gene
Phy A	phytochrome A gene
PMSF	phenylmethylsulfonyl fluoride
rbcS	ribulose-1, 5-bisphosphate carboxylase small subunit gene
SBF-1	silencer binding factor
SV40	simian virus 40
780BP	780 binding protein
780BPE	780 binding protein element
TAF-1	Transcription Activation Factor
T-cyt	*tmr;* T-DNA cytokinin gene
T-DNA	transfer DNA
TGA 1a	(TGACG) tobacco DNA-binding protein
tml	tumor large
tmr	tumor root
tms	tumor shoot
Tris	tris (hydroxymethyl) amino methane
3AF1	AT-rich binding protein
U	units

Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

CHARACTERIZATION OF A PLANT NUCLEAR PROTEIN THAT BINDS A SPECIFIC SEQUENCE OF THE *780* GENE PROMOTER OF T-DNA

By

Eloise Christine Adams

August 1993

Chairman: William B. Gurley
Major Department: Microbiology and Cell Science

A 15 base pair potential *cis* element (780BPE), 5'-TTGAAAAATCAACGC-3', has been identified (-408 to -393) in the promoter of the *780* gene of T-DNA. Using double stranded oligonucleotides synthesized containing sequences homologous to a region in the *780* upstream activator region as a probe and nuclear extracts from commercially obtained cauliflower, DNA-protein interactions were demonstrated using the gel retardation assay. This binding activity is designated the 780 binding protein (780BP). Specific bases required for binding were defined using kinetic competition studies with mutated oligonucleotides, methylation interference assays, $KMnO_4$ interferences assays, and DNase I footprinting. This binding site, which may be a promoter element, was shown to

be different from any plant element previously characterized. With a series of protein purification steps which include ion exchange, gel filtration, heparin affinity, and sequence specific DNA affinity chromatography, the *780* binding protein (780BP) has been purified from cauliflower nuclear extract and characterized. 780BP represents only the second plant factor identified that specifically binds a DNA sequence of a T-DNA promoter. 780BP was characterized as a monomer in solution and has a molecular weight in the range of 48 to 50 kilodaltons. Interaction of 780BP with a mammalian hormone receptor element implies the existence of a DNA binding domain with high homology to the conserved domain 2 of the steroid/thyroid hormone receptor superfamily.

CHAPTER 1

INTRODUCTION

The relationship between *Agrobacterium tumefaciens* and many dicotyledonous plants represents a unique system of naturally evolved genetic engineering. This bacterium has evolved mechanisms to infect and transform eukaryotic plant cells. Upon infection of the dicotyledonous plant cell by *A. tumefaciens*, transfer DNA (T-DNA) moves from the bacterium (tumor inducing plasmid, pTi) to the plant cell and is integrated into the plant genome (Chilton *et al.*, 1977) which results in the development of a crown gall tumor (Chilton *et al.*, 1980). The phenomenon of wound site invasion, integration of T-DNA into the plant genome and crown gall development has been reviewed extensively (Bevan & Chilton, 1982; Nester *et al.*, 1984; Weising *et al.*, 1988).

Genes within the integrated T-DNA are under the control of plant regulatory signals. Of the transcriptional controls identified in T-DNA promoters, only one, the octopine synthase element (*ocs*-element) (Ellis *et al.*, 1987) has been shown to bind plant nuclear proteins. Those purified from tobacco and maize nuclei have been designated as the *ocs*-element transcription factor (OCSTF) (Singh *et al.*, 1989; Tokuhisa *et al.*, 1990). A cDNA clone from a maize root tip library encoding a specific *ocs*-element binding factor

(OCSBF) has been characterized and shown to be a 21 kilodalton (kD) protein with homology to members of the basic leucine zipper family (bZip) of transcription factors (Singh *et al.*, 1990). The fact that T-DNA genes are integrated into the plant genome and are regulated by plant transcription factors dictates that the mode of T-DNA promoter activation will parallel that of endogenous plant promoters.

A novel binding site for a plant nuclear protein(s) is reported here for the *780* gene of T-DNA. Gel retardation assays, methylation interference assays, DNase I footprinting, and competition studies have been used to define this site of protein-DNA interaction. A nuclear protein (or group of proteins) which has been shown to bind specifically to this sequence has been purified from nuclear extracts prepared from cauliflower inflorescences by ion exchange, gel filtration, and sequence-specific DNA affinity chromatography.

Eukaryotic Gene Promoters

Transcriptional control at the molecular level results from interactions between the eukaryotic promoter elements, RNA polymerase, and general and specific transcription factors. Eukaryotic promoter structure and promoter interactions with transcription factors have been studied by many independent groups (Breathnach & Chambon, 1981; Odell *et al.*, 1985; Bruce & Gurley, 1987; Ow *et al.*, 1987; Bucher,

1990). This topic has also been the subject of a number of review articles (Ghosh *et al.*, 1981; McKnight & Kingsbury, 1982; Khoury & Gruss, 1983; Dynan & Tjian, 1985; Serfling *et al.*, 1985; Sassone-Corsi & Borrelli, 1986). A typical animal promoter for RNA polymerase II contains, with few exceptions, the following elements: a TATA box, a cap signal, and a CCAAT box. Plant promoters are very similar in organization with one exception: they generally do not contain the typical CCAAT box motif. Other upstream promoter sequences such as the SV40 core enhancer motif (5'-GTGG$^{A}/_{T}{}^{A}/_{T}{}^{A}/_{T}$-3') (Banerji *et al.*, 1981; Weiher *et al.*, 1983) are being routinely identified in eukaryotic promoters. The TATA motif, which fits the consensus (5'-TATA $^{A}/_{T}A^{A}/_{T}$-3'), is usually located at positions -35 to -25 base pairs (bp) from the transcription initiation site. The TATA box is found in most protein encoding genes and the interaction of this box with the TATA binding protein accurately positions RNA polymerase II over the transcription initiation site (Sawadogo & Roeder, 1985). The CCAAT sequence is commonly found 50 to 100 bp upstream from the start of transcription and is involved in promoter function in animals. Other sequence motifs which appear to be the binding sites of promoter specific transcription factors often function as enhancers and can be found even further upstream.

Much of the original work which defined the role of DNA sequence elements in the regulation of eukaryotic gene expression was conducted in animal viral and cellular gene

systems. For example, an enhancer element was initially identified upstream from the cap site of the early genes of SV40 (Benoist & Chambon, 1981; Jones *et al.*, 1988). The immunoglobulin (Ig) heavy chain gene was the first cellular gene found to contain an enhancer (Banerji *et al.*, 1981; Gillies *et al.*, 1983). The elements contained within these enhancer regions were found to be diverse in sequence as well as in function. Those of the metallothionine-I gene were found to be induced by heavy metals, glucocorticoid hormones, and bacterial lipopolysaccharides. Constitutive elements were also present in the enhancer region of this gene. While some of these enhancer elements were found to be ubiquitous in their activity, others have been shown to be species and tissue specific.

Plant Promoters/DNA Binding Proteins

Plant promoter structure and function in the regulation of plant gene expression is very similar to that of animal promoters (Heidecker & Messing, 1986; Kuhlemeir *et al.*, 1987). For example, in zein genes, a representative group of plant genes, a TATA box and homology to the CCAAT motif and to the animal core enhancer sequences have been identified (Brown *et al.*, 1986). In addition to the TATA box and the putative CCAAT box motif, the cauliflower mosaic virus 35S promoter (CaMV 35S) contains three sequence motifs, *as*-1 (Lam *et al.*, 1989), GATA, and CA (Benfey *et al.*, 1990), which bind

the *trans*-acting factors ASF-1, GATA-1, and CAF, respectively.

Plant gene promoters from a variety of species have been analyzed and found to contain regulatory sequences and sites for DNA-protein interactions. Characterized genes include the small subunit of ribulose-1,5-bisphosphate carboxylase gene (*rbcS*) (Herrera-Estrella, 1984; Morelli *et al.*, 1985; Timko *et al.*, 1985; Fluhr *et al.*, 1986; Green *et al.*, 1987; Kuhlemeir *et al.*, 1987), the chalcone synthase gene (*chs*) (Kaulen *et al.*, 1986; Dron *et al.*, 1988; Schulze-Lefert *et al.*, 1989; Staiger *et al.*, 1989; Lawton *et al.*, 1991; Weisshaar *et al.*, 1991), the phenylalanine ammonia-lyase gene (*pal*) (Lois *et al.*, 1989; Ohl *et al.*, 1990), the small molecular weight heat shock gene *Gmhsp17.5-E* (Czarnecka *et al.*,1989; Barros *et al.*, 1992), and the alcohol dehydrogenase gene (*Adh1*) (Ferl & Nick, 1987; Walker *et al.*, 1987; Ferl & Laughner, 1989; DeLisle & Ferl, 1990). The regulatory sites are of varying sizes and complexities. The 6 bp *as*-1 sequence in the CaMV 35S promoter (Lam *et al.*, 1989) is among the smaller sites and the 280 bp upstream sequence from *rbcS* (Kuhlemeir *et al.*, 1987) is among the larger sites. The larger upstream sequences are generally found to contain a number of smaller regulatory elements. For example, within the 280 bp sequence of *rbcS* short conserved sequences labeled box I, box II, and box III were identified and shown to be light responsive elements (LRE) (Kuhlemeir *et al.*, 1987). A 100 bp upstream region of the *chs* gene promoter was

found to contain two smaller sequences that directly influence gene expression (Kaulen *et al.*, 1986; Schulze-Lefert *et al.*, 1989). The soybean heat shock gene promoter (*Gmhsp17.5-E*) has two heat shock elements HSE-1 (Pelham, 1982) (-72 to -49), HSE-2 (-103 to -81) and two AT-rich regions (-159 to -120) (-234 to -201) (Czarnecka *et al.*, 1989). In an extensive mutagenesis analysis of HSE-1, the optimum 5 bp repeat sequence was defined as either 5'-aGAAg-3' or 5'-cTTCt-3' (Barros *et al.*, 1992). The *Adh1* promoter is also a complex promoter with a number of regulatory elements controlling not only aerobically induced expression, but constitutive expression and organ specific expression as well (Howard *et al.*, 1987). There is an anaerobic regulatory element (ARE) between positions -140 and -99 (Walker *et al.*, 1987) as well as G-box homology at positions -195 to -184 and at positions -122 to -114 (Ferl & Nick, 1987).

In addition to positive elements, negative elements have also been characterized as illustrated by a 153 bp region from the *chs15* gene promoter which was analyzed and functionally characterized as a transcriptional silencer (Dron *et al.*, 1989). This region was later shown to contain three regions, box 1, box 2, and box 3 which are the binding sites for the nuclear binding silencer factor-1 (SBF-1) (Lawton *et al.*, 1991).

Gel retardation assays, *in vivo* dimethyl sulfate footprinting, *in vitro* methylation interference assays, and DNase I footprinting have been used to identify a variety of

DNA-binding proteins that interact with specific plant promoter sequences. Though a number of these DNA binding proteins have been cloned, most still can not be definitely classified as transcription factors since direct involvement in transcriptional regulation has not been demonstrated.

The plant transcription factors and DNA binding proteins cloned to date are almost exclusively members of the bZip group (Landschulz *et al.*, 1988; Vinson *et al.*, 1989). Plant bZip proteins appear to bind two distinct DNA motifs with a small number able to interact with both types of elements. The two DNA classes of elements, as studied in detail by the Cashmore and the Chua groups, are the G-box core hexanucleotide (5'-CACGTG-3') motif (Williams *et al.*, 1992) and the 5'-tgACGT/C-3' motif (Schindler *et al.*, 1992b).

The core hexanucleotide (5'-CACGTG-3') has been identified in a wide variety of plant promoters including *rbcS* genes from pea, tobacco, soybean, and *Arabidopsis* (Fluhr *et al.*, 1986; Grandbastien *et al.*, 1986; Guiliano, 1988). Although the original report (Guiliano, 1988) defined a 12 bp sequence, 5'-GACACGTGGC-3', to be sufficient for binding the G-box factor (GBF), it is the core hexanucleotide embedded within this sequence that has become identified as the G-box element. Wheat nuclear protein EmBP-1 has been shown to bind G-box-like sequences found in the ABA-responsive element of the wheat EM promoter and in the wheat histone 3 promoter (Guiltinan *et al.*, 1990). HBP-1a, also known as HBP-1, binds strongly to the wheat histone 3 promoter sequence (Tabata *et*

al., 1989; 1991). A tobacco bZip protein, TAF-1, has been shown to bind a G-box-like motif and to activate transcription (Oeda *et al.*, 1991). In addition nuclear protein extracts from tomato leaves and *Arabidopsis* seedlings (Guiliano, 1988), tobacco seedlings (Staiger *et al.*, 1989; 1991), and maize suspension culture cells (DeLisle & Ferl, 1990) have been shown to bind the G-box. Numerous G-box binding proteins have been cloned and include, wheat EmBP-1 (Guiltinan *et al.*, 1990), tobacco TAF-1 (Oeda *et al.*, 1991), parsley CPRF-1,2,3 (Weisshaar *et al.*, 1991), HBP-1a and HBP-1b (Tabata *et al.*, 1991), GBF 1, GBF 2, and GBF 3 (Schindler *et al.*, 1992a).

A general analysis of the G-box motif, the two 5' bases, and the two 3' bases flanking the hexamer has resulted in the grouping of the G-box elements into two classes (Williams *et al.*, 1992). In the study by Williams *et al.* (1992), sixteen oligonucleotides were synthesized which contained an intact hexameric core and all possible combinations of nucleotides at the two bases flanking both the 5' and the 3' ends. A mutant hexamer core (5'-CAATTG-3') with flanking AT sequences was also synthesized and used as a probe in gel retardation assays with cauliflower nuclear extract. Class I G-box elements form the type A protein-DNA complex. This complex is characterized by a large diffuse band which upon shorter exposure times is seen to have three components (A1, A2, and A3). Class II G-box elements form the type B protein-DNA complex characterized by three major protein-DNA subcomplexes

(B1, B2, B3). The perfect palindromic 8mer or 10mer fell into class I. All G-box-protein interactions described to date have been classified according to the predicted binding pattern (Williams *et al.*, 1992). One might infer from these studies that in a cell expressing a limiting concentration of type A binding proteins and abundant type B binding proteins, genes whose promoters have class II binding sites would be preferentially expressed over those with promoters containing class I sites. This phenomenon of differential affinity between members of a class of *trans*-acting factors has been postulated to play an important role in the regulation of gene expression (Williams *et al.*, 1992).

The ACGT core, which is a component of the G-box element and the 5'-TGACGT/C-3' motif, has also been identified in the promoter of the octopine synthase gene as the 16 bp palindrome *ocs*-element (5'-ACGTAAGCGCTTACGT-3') (Ellis *et al.*, 1987; Singh *et al.*, 1989). The 20 bp consensus *ocs*-element (5'-TGACGT/CAAGC/GG/AA/CTG/TACGT/CA/CA/C-3') defined by Bouchez and colleagues also contains the ACGT core (Bouchez *et al.*, 1989). Bouchez and colleagues also identified six opine synthase gene promoters of T-DNA (in addition to *ocs*) from Ti and Ri plasmids and three plant viral gene promoters that contain the *ocs*-element. The nuclear protein OCSTF (Singh *et al.*, 1989; Tokuhisa *et al.*, 1990) and the cloned OCSBF (Singh *et al.*, 1990) interact with the *ocs*-element to form two protein-DNA complexes. Whether

or not OCSTF and OCSBF are the same protein has yet to be determined.

DNase I footprinting and gel retardation assays were used to demonstrate that the protein factor ASF-1 purified from pea whole cell extract and nuclear extracts of green pea seedlings and tobacco leaves is able to bind the *as*-1 sequence (5'-TGACG-3') in the CaMV promoter (located between -85 to -58) (Lam *et al.*, 1989). A tobacco cDNA clone was isolated that encodes a DNA-binding protein that binds *as*-1 and was named TGA1a (Katagiri *et al.*, 1989). Because both proteins specifically interact with the same sequence, and mutations within this sequence abolish the binding of both and cause a loss of *as-1* function, the cloned TGA1a protein and the purified ASF-1 from plant cells are believed to be the same (Katagiri *et al.*, 1989; Lam *et al.*, 1989).

Proteins that bind ACGT core motifs and show only a two amino acid difference in the basic DNA binding domain can exhibit different DNA binding site preferences. This was demonstrated when amino acid sequences in the basic region of HBP-1a (Tabata *et al.*, 1991) were compared with those of GBF-1 and found to differ by only two residues (Schindler *et al.*, 1992b). GBF-1 binds the perfect palindromic G-box of the parsley *chs* promoter with equal affinity with the 5'-tgACGTGG-3' motif of the wheat histone 3 promoter. HBP-1a, on the other hand, binds to the wheat histone 3 hexamer sequence with high affinity and binds to the parsley *chs* promoter G-box with low affinity (Tabata *et al.*, 1991). This

suggested that regions outside of the DNA binding domain of the protein contribute to DNA binding specificity. This hypothesis was supported when Schindler *et al.* (1992a) showed that converting the GBF-1 DNA binding domain to a HBP-1a DNA binding domain failed to alter the DNA binding properties of GBF-1.

Diversity within a family of DNA sequence motifs can greatly affect binding affinity of a regulatory factor. This is evident in the GBF 1 binding site analysis by Schindler and colleagues (Schindler *et al.*, 1992b). The high affinity binding site of GBF 1 was shown to be the palindromic G-box (5'-CCACGTGG-3') (Schindler *et al.*, 1992b). Several other high affinity DNA binding sites, which all shared a 5'-ACGTG-3' core, were identified using the random binding site selection assay. High affinity sequences with specific permutations of the bases (C,A,T,G) at the two most 5' positions outside of the ACGTG core are CCACGTG, ACACGTG, TGACGTG, CTACGTG, and TTACGTG. The sequence 5'-TGACGTXX-3' is recognized by GBF 1 only if the two 3' bp are GG or GT.

In vivo dimethyl sulfate footprinting experiments were used to define sequences within the promoters of the maize *Adh1* gene (Ferl & Nick, 1987) and the *Arabidopsis Adh* gene (Ferl & Laughner, 1989) as binding sites of regulatory molecules. Four regions were found within these promoters to share unique homology. One of these regions was designated the 4C box and three designated the GTGG box. An *Arabidopsis* G-box binding factor from whole cell extracts prepared from

suspension cultures and leaves has also been characterized for these G-box sequences in the *Arabidopsis Adh* gene promoter (DeLisle & Ferl, 1990). A synthetic oligonucleotide including a specific 17 bp site from the ARE (encompassing the -130 region) in maize was used in gel retardation experiments with whole cell extracts made from maize cell suspension cultures to define the ARE binding factor (ARF-B_2) (Ferl, 1990).

The pea nuclear factor GT-1 when cloned and analyzed will most likely not be a member of the bZip class of DNA binding proteins because the cognate binding sequences do not contain the ACGT core. GT-1 was identified as the factor interacting with light responsive elements box II, 5'-GTGTGGTTAATATG-3' (-151 to -138), and to box III, 5'-ATCATTTTCACT-3' (-125 to -114) of the *rbcS-3A* promoter (Green *et al.*, 1987; 1988). It also binds the redundant sequences box II* and box III* further upstream. Through substitution mutational analyses, six bases within box II (GGTTAA) were found to be critical for the binding of GT-1 (Green *et al.*, 1988). GT-1 is present in the leaves of both light-grown and dark-grown plants. It is not clear at this time whether this binding results in positive or negative regulation since the positive and the negative elements overlap.

A protein with binding sequence preference similar to that of GT-1 has been cloned from rice seedlings. This protein called GT-2, while having no obvious sequence homology to other proteins in the databases, contains a

proline- and glutamine-rich domain, separate basic and acidic regions, and a segment with the potential to form a helix-loop-helix structure (Dehesh *et al.*, 1990). Sequence preference for cloned GT-2 is 5'-GCGGTAATT-3' (-228 to -219) in the rice *phytochrome A* (*phy A*) gene promoter. This protein binds to the box II element (5'-GTGTGGTTAAT-3') in the pea *rbcS* promoter, albeit with two orders of magnitude lower affinity than to the rice *phyA* sequence. The binding of nuclear factor GT-1 to box II and to box III with high affinity potentially distinguishes nuclear factor GT-1 from cloned factor GT-2.

Another group of plant proteins, AT factors, have been shown to bind to AT-rich sequences present in numerous plant promoters. AT-rich sequences have been reported in plant genes that are developmentally regulated (Jofuku *et al.*, 1987; Jensen *et al.*, 1988; Bustos, 1989; Jordano *et al.*, 1989; Rocha-Sosa *et al.*, 1989; Jacobsen *et al.*, 1990), light-regulated (Datta, 1989), and stress-regulated (Baumann, 1987; Czarnecka *et al.*, 1989; Harrison, 1991a; 1991b). The pea nuclear factor AT-1 binds to a specific AT-rich sequence within the promoters of nuclear genes encoding the *rbcS* and the polypeptide component of the light-harvesting chlorophyll a/b protein (*cab*) complex. This sequence is termed the AT-1 box. Some of these AT-rich sequences have been shown to be regulatory in nature and shown to bind nuclear proteins (Czarnecka *et al.*, 1989). However, only one AT factor, 3AF1, has been cloned to date (Lam *et al.*, 1990). An

oligonucleotide containing a tetramer of the AT-rich Box VI (-51 to-31) from the promoter of the *rbcS-3A* gene was used to screen an amplified tobacco leaf expression library. A partial cDNA clone was obtained, analyzed, and found to contain a repeat of about 94 amino acids with histidines and cysteines reminiscent of the zinc finger DNA binding motif (Struhl, 1989). Since this region does not conform exactly to the well characterized zinc finger motif, classification of this clone as a zinc finger protein was withheld.

As evident from the previous discussion, plant promoters contain a diverse group of regulatory elements which interact with a variety of DNA binding proteins. The promoters of T-DNA genes expressed in plants have also been shown to contain an array of regulatory elements and could also prove to interact with numerous plant DNA binding proteins.

T-DNA Gene Promoters

Studies utilizing the soil bacterium *A. tumefaciens* have given insights into many fundamental processes of plant biology (Klee *et al.*, 1987). This bacterium can be used as a vector for plant transformations; thus, the ability to create transgenic plants presents a variety of research opportunities. For example, this technology has been used as a tool for the engineering of plant resistance to herbicides and viruses, for studying tumorigenesis in plants, and for studying the regulation of plant gene expression.

Ti plasmids are classified according to the type of opines their encoded genes induce. Opines are sugar-amino acid derivatives that the *Agrobacterium* metabolizes as a carbon and nitrogen source (Holsters *et al.*, 1982). The two most commonly studied plasmids are those that encode nopaline synthase and octopine synthase. T-DNA from the nopaline plasmid is transferred as one contiguous 22 kb fragment. However, T-DNA from the octopine type plasmid is transferred either as one distinct fragment (23 kb), or as multiples of T-left DNA (13 kb), and T-right DNA (7.8 kb) (Thomashow *et al.*, 1980; Barker *et al.*, 1983). The presence of two imperfect 24 bp repeats on the left and right borders, and two similar 24 bp repeats within the T-DNA region divide the octopine type plasmid (pTi15955) into three distinct domains: T-left (T_L), 13175 bp; T-center (T_C), 1816 bp; and T-right (T_R), 7883 bp (Barker *et al.*, 1983).

When the complete nucleotide sequence of pTi15955 was determined, 26 open reading frames were identified (Barker *et al.*, 1983), but only 13 transcripts have been reported to date (Fig. 1-1) (Willmitzer *et al.*, 1981; 1982b; Winter *et al.*, 1984). Eight of these genes (*5*, *7*, *tms 1* and *2*, *tmr*, *ons*, *tml*, and *ocs*) are located on T_L, and the remaining five (*780* (*4'*), 1040 (*3'*), *mas* (*2'* and *1'*), and *ags* (*0'*)) are located on T_R. Gene *7* encodes mRNA of unknown function. Gene *5* modulates the auxin response in plants (Korber *et al.*, 1991). The *tms1* and *tms2* genes encode enzymes of a two-step pathway for the biosynthesis of auxins, and the *tmr* gene

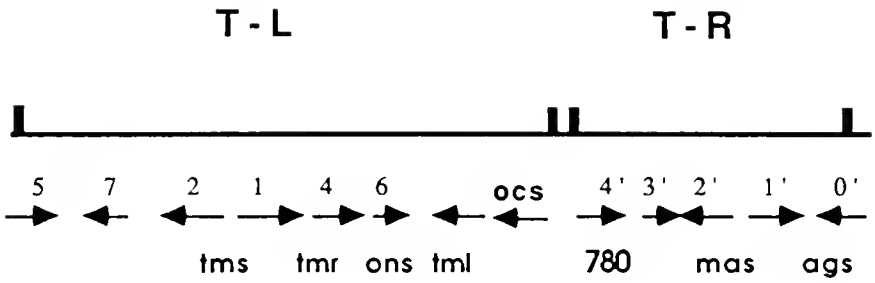


Figure 1-1. Map of octopine T-DNA [Leisner, 1989]. Rectangles denote the T-DNA borders and arrows indicate the direction of transcription. A description of individual genes is given in the text.

encodes an enzyme for the biosynthesis of cytokinins. The *ons* gene encodes a protein involved in octopine and nopaline secretions (Messens *et al.*, 1985). The *tml* gene regulates tumor size in some plant species. The *ocs* gene encodes an enzyme that synthesizes octopine (Hack & Kemp, 1977). The *780* gene and the *1040* gene (Winter *et al.*, 1984) encode mRNAs of unknown functions. The *mas* genes encode enzymes of a two-step pathway for the synthesis of mannopine, and the *ags* gene encodes an enzymes that cyclizes mannopine to yield agropine (Ellis *et al.*, 1984; Komro *et al.*, 1985). The genes within the T-DNA from *A. tumefaciens* are transcribed by RNA polymerase II (Willmitzer *et al.*, 1981) and encode a set of well defined polyadenylated transcripts (Koncz *et al.*, 1983). Because the promoters of these genes possess many eukaryotic characteristics and are under the control of plant regulatory factors, they represent good models for studying protein-DNA interactions involved in the control of plant gene expression.

The *nos* (Koncz *et al.*, 1983; Shaw *et al.*, 1984; An *et al.*, 1986; Ebert *et al.*, 1987; Mitra & An 1989), *mas* (Velten *et al.*, 1984; DiRita & Gelvin, 1987; Langridge *et al.*, 1989; Leung *et al.*, 1991; Fox *et al.*, 1992), *ocs* (DeGreve *et al.*, 1983; Knocz *et al.*, 1983; Ellis *et al.*, 1987a; 1987b; Leisner & Gelvin, 1988), *ags* (Ellis *et al.*, 1984; Bandyopadhyay *et al.*, 1989), *T-cyt* (dePater *et al.*, 1987a; 1987b; Strabala *et al.*, 1993) and the *780* (Bruce & Gurley, 1987; Bruce *et al.*, 1988) promoters have been dissected, analyzed, and well

characterized. They have all been shown to be composed of multiple regulatory regions and therefore have potential for being the site of protein-DNA interactions.

The *ocs*-element has been identified in the *nos* promoter (-130 to -101) (An *et al.*, 1986) and was shown to be essential for promoter activity. The *nos* promoter, originally thought only to be constitutively expressed in plants (An *et al.*, 1986), has subsequently been shown to be organ specific and developmentally controlled in transgenic tobacco plants (An *et al.*, 1988). It was also shown that this promoter is wound inducible in a variety of vegetative and reproductive organs (An *et al.*, 1990). The wound response mediated by the *ocs*-element of *nos* (*nos* element) seems to be enhanced by auxin.

The *mas* promoter represents a dual promoter system with two overlapping promoters of opposite polarity which have regions that are distinct and others that are shared (Velten *et al.*, 1984; DiRita & Gelvin, 1987; Langridge *et al.*, 1989; Leung *et al.*, 1991). At least three *mas* elements (Fox *et al.*, 1992) clearly related to the *ocs*-element overlap many of the previously reported regulatory elements of Leung and colleagues (Leung *et al.*, 1991). The *mas* promoter has been used a number of times in constructs with heterologous promoters in an effort to boost transcription levels (Harpster *et al.*, 1988; Langridge *et al.*, 1989; Teeri *et al.*, 1989; Sanger *et al.*, 1990; Leung *et al.*, 1991). Higher expression patterns have been observed in roots with the *mas*

promoter. Expression was also found to be wound inducible in leaf and stem tissue and to be auxin and cytokinin responsive in normal and tumorous plant tissues.

Deletion and substitution mutagenesis experiments with promoter sequences of the *ags* gene in sunflower crown gall tumors have revealed five regulatory regions in addition to TATA (Bandyopadhyay *et al.*, 1989). This promoter also has a region with sequence homology to the *ocs*-element (Barker *et al.*, 1983; Bouchez *et al.*, 1989); however, the role of this sequence has yet to be demonstrated.

The *T-cyt* (*tmr*) promoter has been analyzed and found to contain upstream regulatory sequences (dePater *et al.*, 1987a; 1987b). The upstream region between -442 and -408 was shown to be responsible for maximum expression in roots, but not in other organs nor in a tobacco suspension-cultured cell line (Strabala *et al.*, 1993). The effect of the -185 to -129 deletion on the expression of *T-cyt* is in dispute. In tumors, deletion of this region resulted in significant decrease in expression from the *T-cyt* promoter (dePater *et al.*, 1987) In tobacco cell suspension cultures, no alteration in promoter strength was observed when this region was deleted (Strabala *et al.*, 1993). This may prove not to be a discrepancy at all, but simply a case of tissue specific expression.

Octopine Synthase Gene Promoter

The *ocs* gene is located on the extreme right border of T_L from the octopine type Ti plasmid. It codes for the enzyme octopine synthase (Schroder *et al.*, 1981) which catalyzes the reductive condensation between pyruvate and arginine, lysine, histidine, or ornithine to yield, respectively, octopine, lysopine, histopine, or octopinic acid (Hack & Kemp, 1977). Sequence analyses indicate the presence of a TATA box, translation from AUG, a 3' polyadenylation recognition signal, no introns, and a G+C content similar to that of many plant genes. This gene and its upstream regulatory sequences is one of the most well studied of the T-DNA genes. When promoter mutations of the *ocs* gene were introduced in crown gall tumors to study their effects on expression, deletions upstream of position -294 were found not to interfere with expression (Koncz *et al.*, 1983). However, deletions upstream of -170 greatly reduced gene expression. The promoter region was thus hypothesized to be 295 bp in length with a 125 bp control region.

A 176 bp (-296 to -116) sequence of the *ocs* promoter was shown to activate the maize *Adh1* promoter in transgenic tobacco plants (Ellis *et al.*, 1987b). Because this element acted independently of orientation, these authors began using the term "enhancer" in reference to the enhancer element of animal genes. However, the ability of this fragment to act

3' and at great distances was not tested by the same authors. Transient gene expression using deletion mutants and synthetic oligonucleotides in maize protoplasts was used to identify a 16 bp palindrome sequence, 5'-ACGTAAGCGCTTACGT-3' (-193 to -178), as a major regulatory component of the 176 bp fragment (Ellis *et al.*, 1987a). Chimeric maize *Adh1/chloramphenicol acetyltransferase* genes with a variety of promoter mutant constructs were electroporated into maize protoplasts and assayed for CAT activity. When the *ocs*-element was positioned at -163 bp with respect to the transcription start site, *Adh1* promoter activity was increased 200-fold over basal levels. Insertions of 275 to 1000 bases between the *ocs*-element and the transcription start site resulted in a dramatic decrease in *Adh1* promoter activity. Thus, the influence of the *ocs*-element diminishes with increasing distance from the *Adh1* promoter. The strength of the element was also diminished when it was placed in the 3' position. As a consequence, this activator region is referred to as having enhancer-like properties. Animal enhancers can act at great distances and 3' to the promoter proper (Banerji *et al.*, 1981; Fromm & Berg, 1983).

The *ocs*-element is necessary for *ocs* gene expression in transformed tobacco calli (Leisner & Gelvin, 1988). In this system, Leisner and Gelvin show that the element functions independently of orientation when placed upstream of the *ocs* gene, which supports the previous finding. However, a minor discrepancy is observed in their report that the *ocs*-element

does not function in either orientation when placed downstream of the gene, nor does it activate its promoter when placed 608 bp away. A negative element is hypothesized to be located between -249 and -222 since promoter activity is inhibited when these sequences are present in the same construct with the *ocs* palindrome.

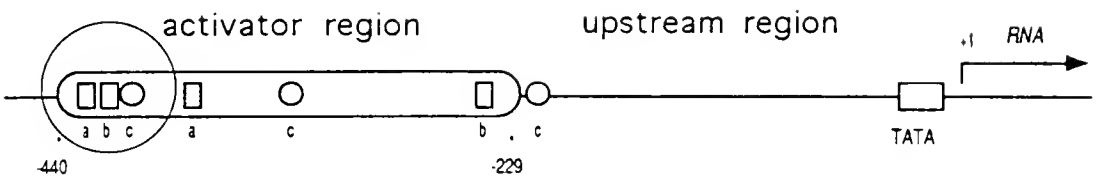
Gene expression under the control of the *ocs* upstream activator region (contains the *ocs*-element) was shown to be constitutive in leaves, roots, and stems of transgenic plants (Otten *et al.*, 1981). More recently, the *ocs*-element has been found to be tissue specific and developmentally regulated (Fromm *et al.*, 1989). Using glucuronidase synthase (GUS) reporter gene constructs, the *ocs*-element was shown to confer expression confined to the root tip and shoot apex of transgenic tobacco plants. The apparent discrepancy in constitutive verses organ specific expression may not be a real discrepancy at all. Otten *et al.* used the full upstream promoter element (Otten *et al.*, 1981) in which the expression pattern observed may have been the result of the action of a combination of *cis* elements located elsewhere in the promoter. The Fromm group used two smaller promoter sequences, a 59 bp fragment and a 21 bp fragment from the *ocs* promoter, both of which contained the *ocs*-element (Fromm *et al.*, 1989). It was also shown that the 16 bp palindrome directs cell specific expression of the *ocs* promoter within the leaves, stems, and roots of transgenic plants (Kononowicz *et al.*, 1992).

The existence of sequences homologous to one half of the *ocs* palindrome within the 5'-flanking regions of plant genes such as the cauliflower mosaic virus 35S promoter (Odell *et al.*, 1985), the potato protease inhibitor II gene (Thornburg *et al.*, 1987), and the soybean lectin (*Le1*) gene (Vodkin *et al.*, 1983) suggests not only the existence of homologous T-DNA promoter elements in plant gene promoters, but also suggests that modifications have occurred for purposes of regulatory control.

780 gene promoter

The *780* gene, so called because of its approximate 780 bp transcript (Winter *et al.*, 1984), is the major focus of the research presented here in this dissertation. It is located on the extreme left of T_R-DNA of the octopine Ti plasmid (Fig. 1-1) (Karcher *et al.*, 1984; Winter *et al.*, 1984). The promoter region has been characterized by a series of 5' and internal deletions and substitutions (Bruce & Gurley, 1987). The effect on transcriptional activity from both the major and minor start sites was quantitated from sunflower crown gall tumors. Three functional domains, an activator, an upstream promoter region, and TATA box, were defined (Fig. 1-2). Within the activator(-440 to -229) three direct repeats (*a*, *b*, *c*) were identified (Bruce & Gurley, 1987). A cluster of these repeats is found between -427 to -396 and may be involved in the sharp reduction of transcription when this region is deleted. The upstream

780 gene of T-right



repeats:

a = TCCTTTCGAC

b = CACGGA

c = TTGAAAA

Figure 1-2. The *780* gene promoter. The *780* activator region (-440 to -229) in context with the upstream region (-229 to -37) and the TATA element (-37). Circled region denotes clustered repetitive sequences a, b, and c.

region (-229 to -37) between the activator and TATA has been designated the upstream promoter element (Bruce & Gurley, 1987). Its effect on transcription was observed when internal deletions (-76 to -74 and -112 to -98) resulted in dramatic reductions in transcriptional activity. Also within this region, there is homology to the CCAAT box consensus (-120 in the minor promoter, and -60 in the major promoter) seen in the upstream elements of animal genes (Dynan & Tjian, 1985). In addition to various upstream elements, the TATA element was also shown to be required for transcriptional activity.

The *780* activator possess some of the characteristics typical of an animal enhancer evidenced by its ability to stimulate transcription in a bi-directional manner over a relatively large distance (650 bases) from TATA in the absence of other upstream elements (Bruce *et al.*, 1988). However, it did not stimulate transcription when positioned 3' of the coding region of the *780* gene (200 bp downstream from poly(A) site). Thus, the name "*780* activator" remains and the term "enhancer-like" is used in reference to the activity of this element. It may be possible that the *780* activator(-476 to -229) is indeed an enhancer. Is it possible that when the 3' construct was generated, not enough of the core promoter ("promoter proper") was left intact. The test gene (-37 to +926) included only six bases 5' of TATA. There may be sequences important for transcription of this gene located in this deleted region since small internal

deletions at -76 and at -112 result in reduced transcriptional activity (Bruce & Gurley, 1987). When the downstream construct was tested, it only had a TATA and a 3' enhancer. In order to definitely classify the 780 activator as an enhancer, a repeat of this experiment is suggested using a construct that contains more intact core promoter sequences.

Plant DNA binding proteins (putative transcription factors) along with their cognate promoter binding sequences have been presented. T-DNA promoters which have been shown to be transcribed by RNA polymerase II, and to be regulated by plant nuclear factors, have also been discussed. The cloning of a plant transcription factor (OCSBF) that regulates a specific T-DNA promoter suggests that there are more. The identification of another such factor has begun with the following report of a novel promoter site of protein-DNA interactions within the promoter of the *780* gene of T-DNA. The process continues with the purification and characterization of a DNA binding protein(s) from cauliflower nuclear extracts that binds this specific site in the *780* gene promoter.

CHAPTER 2

CHARACTERIZATION OF PROTEIN:DNA INTERACTION

Introduction

Agrobacterium tumefaciens induces abnormal proliferation of plant cells resulting in crown gall tumors (for review see Nester *et al.*, 1984). Crown gall tumors can be characterized as undifferentiated, rapidly dividing cells. Tumor induction results from the enlistment of plant transcription factors required for expression of the cytokinin gene, auxin genes, and the genes responsible for the production of opines. T-DNA promoter function has been characterized in meristematic tissue (undifferentiated rapidly dividing cells) (Leisner & Gelvin, 1989). The regulatory element from the *ocs* gene has been found to direct tissue specific expression in root tips of young transgenic tobacco seedlings, and in the shoot apex of older seedlings (Fromm *et al.*, 1989). The highest level of OCSBF-1 expression was found in dividing and differentiating cells of maize, such as the basal section of developing leaves and in roots and shoots of young plants (Singh *et al.*, 1990). Thus, the choice of the cauliflower inflorescence as the source of nuclear extract was not only one of convenience, but was the most reasonable as well,

given the fact that it is an abundant source of meristematic tissue.

One objective of this work is to identify a region in the *780* gene promoter (Fig. 2-1) that interacts with a sequence specific DNA-binding protein. In the deletion mutagenesis analysis of the *780* gene promoter by Bruce and Gurley (1987), removal of a 31 bp region between -427 and -396 resulted in a 45% decrease in the relative transcription level which is indicative of the presence of an upstream regulatory element within these 31 bp. Within this region there are three short repetitive sequences that are clustered together and repeated two to three times within the *780* gene activator. Short repeats of this nature have been shown not only to be sites of protein-DNA interactions in enhancers and upstream elements, but also shown to be regulatory elements of both plant and animal genes (Oeda *et al.*, 1991; Cooney, 1992; Schindler *et al.*, 1992; Williams *et al.*, 1992). Initial studies focused on these repeat sequences; however, this approach was not productive. While these previously identified repeats remain of great interest, the present study focuses on the inverted repeat (5'-TGAA----TTCA-3') also found within this 31 bp region (Fig. 2-2). Specific classes of regulatory factors have been shown to interact with inverted repeats. The role of this inverted repeat and flanking bases in protein-DNA interactions, and perhaps transcriptional regulation, is explored here. The boundaries and the specific bases that are required for DNA-protein

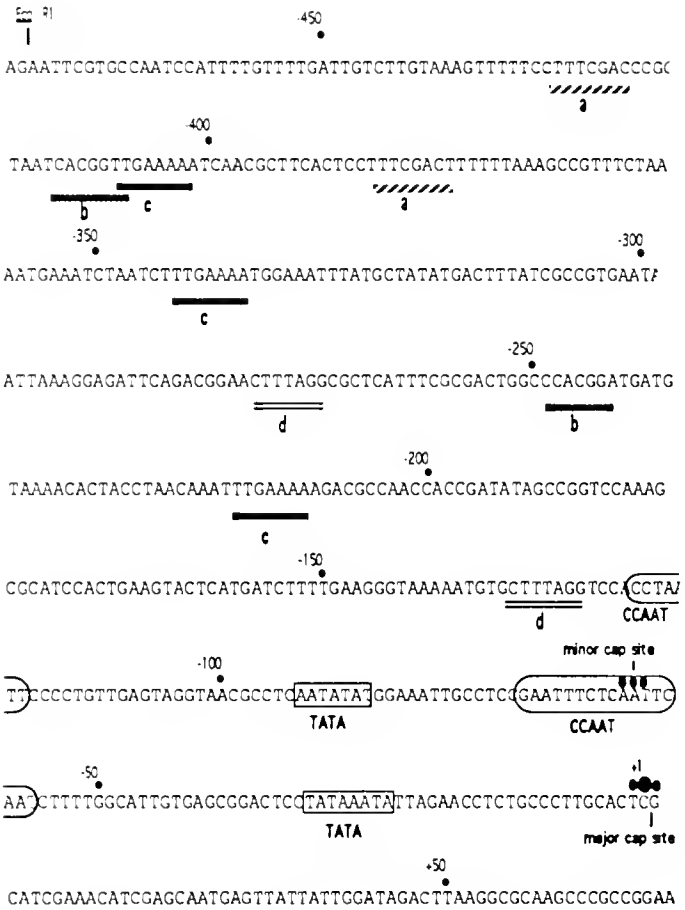


Figure 2-1. Sequences of the *780* gene promoter [Bruce, 1987]. The short repeats *a*, *b*, *c*, and *d* are indicated. Protein-DNA interactions in the region between -410 and -380 are analyzed in this work.

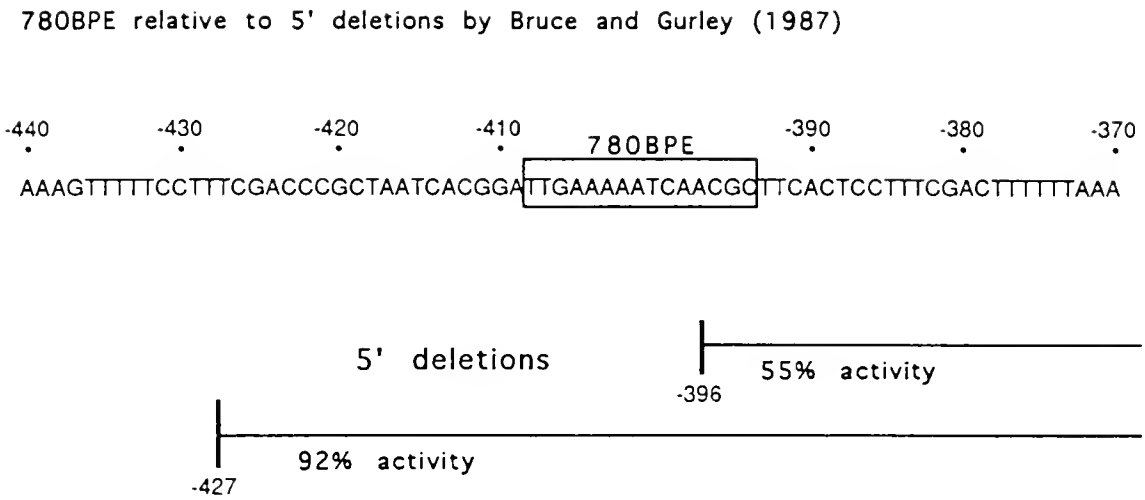


Figure 2-2. 780BPE relative to 5' deletions analyzed by Bruce and Gurley (1987). A double stranded oligonucleotide (-410 to -383) was synthesized and used as a probe for gel retardation assays and footprint analyses.

interaction with a factor in cauliflower nuclear extracts were determined.

Gel retardation, methylation interference assays, $KMnO_4$ interference assays, DNase I footprinting, and competition studies have been used to characterize DNA-protein interactions *in vitro*. The observation that free DNA could be separated from DNA-protein complexes on the basis of their electrophoretic mobilities in polyacrylamide gels lead to the development of the gel retardation assay (Fried & Crothers, 1981; Garner & Revzin, 1981). The gel retardation assay has been used in a quantitative manner (Garner & Revzin, 1981) in kinetic competition studies to define a specific binding sequence. Chemical modification of DNA has been exploited in various applications to identify specific sites of a DNA fragment that are required for protein binding. The modification of guanine at the N-7 position with dimethyl sulfate and the subsequent cleavage by piperidine at the modified base (Maxam, 1980) is a widely used method for the identification of significant G residues in the formation of a DNA-protein complex. Potassium permanganate, historically known for its ability to oxidize double bonds, has been used in reactions in which it preferentially modifies thymine residues (Rubin *et al.*, 1980). Significant T residues for the formation of a DNA-protein complex have been identified using this chemical modification method. DNase I footprinting is a technique that combines the Maxam-Gilbert DNA-sequencing method and the DNase protected fragment

technique (Galas & Schmitz, 1978). This technique has been used extensively for the characterization of DNA-protein interactions in both plant and animal systems (Sawadago & Roeder, 1985; Brenowitz *et al.*, 1986; Green *et al.*, 1987; Guiliano, 1988). Double stranded DNA is partially degraded by DNase in the presence and absence of a binding protein. These single end-labeled fragments are then visualized by electrophoresis and autoradiography along side a base specific reaction product of the Maxam-Gilbert sequencing method. This results in the identification of a protected area, "the footprint" of the binding protein on the DNA fragment.

Materials and Methods

Preparation of nuclear extract.

Nuclear extracts were prepared by modifications of a method described by Wu *et al.* (1987). This entire procedure was performed in the cold room at temperatures between 4-6 °C. Florets from commercially acquired cauliflower heads (ca. 1000 g) were cut, and then crushed with mortar and pestle in 1200 ml of buffer I (10 mM HEPES, pH 7.9; 0.3 M sucrose; 0.1 mM ethylene glycol-*bis* (β-amino ethyl ether)-N; N, N',N'-tetraacetic acid (EGTA), pH 7.9; 1.5 mM $MgCl_2$; 10 mM KCl; 3.0 mM dithiothreitol (DTT); 0.01% NP-40 detergent). The following protease inhibitors were included: 0.5 mM phenlymethylsulfonyl fluoride (PMSF), 4 mM N-ethylmaleimide

(NEM), 1 mg/ml pepstatin A (Sigma), and 0.5 mg/ml leupeptin (Sigma). The tissue was then divided into four samples and each sample ground in 300 ml of buffer (total volume) with a Tekmar Tissuemizer for 3 min at 75% of full power. The homogenate was filtered through miracloth (Calbiochem) wrapped in cheesecloth. Cells were pelleted by centrifugation at 5 K rpm for 10 min at 4 °C. The pellet was resuspened in 15 ml of buffer I, then cells disrupted with 12 strokes of the Dounce hand homogenizer (pestle B). All four samples were pooled and treated as one sample for the remainder of the extraction protocol. Nuclei were pelleted at 5 K rpm for 10 min at 4 °C. The pellet was resuspened to a total of 16-20 ml in buffer II (10 mM HEPES, pH 7.9; 10% glycerol; 0.1 mM EGTA, pH 7.9; 1.5 mM $MgCl_2$; 0.45 M KCl; 3.0 mM DTT; 0.01% NP-40; 0.5 mM PMSF; 4 mM NEM; 1 mg/ml pepstatin A; and 0.5 mg/ml leupeptin) and extraction allowed to proceed for 30 min by stirring on ice in the cold room. The nuclear extract was clarified with a one hr spin in the ultracentrifuge (Bechman SW55 rotor) at a speed of 32 K rpm (100,000 X g). The nuclear extract was then aliquoted, frozen in liquid nitrogen and stored at -80 °C. The protein concentration was later determined using a Bradford Protein Assay kit from BioRad.

Gel retardation assay.

DNA binding activity was determined by gel retardation assays modified from Singh *et al.* (1986). Crude nuclear

extract (1-3 µg) was incubated at room temperature for 30 min with 1 µg of poly(dI-dC)-poly(dI-dC), 1 µg yeast tRNA (Sigma), and 5 x 10^4 cpm of oligonucleotide probe 3'-labeled by the fill-end reaction (Sambrook *et al.*, 1989) in 25 µl of binding buffer (10 mM HEPES pH 7.9; 10% glycerol; 0.1 mM EGTA, pH 7.9; 100 mM KCl; 1.5 mM $MgCl_2$; 0.45 M KCl; 3.0 mM DTT; 0.01% NP-40).

The binding reaction was loaded on an 8% polyacrylamide gel (30:1, acrylamide to bis-acrylamide) and electrophoresed in a 10 mM Tris, 10 mM boric acid, 30 mM EDTA buffer (TBE) at 200 volts for 2 hrs. The gel was dried under vacuum on a 3MM Whatman filter using a heated slab drier and exposed overnight to X-ray film (Kodak, PDB-1) in the presence of an intensifier screen at -80 °C. For quantitation purposes, some gels were exposed to a PhosphorImager screen.

Kinetic competition studies

A series of mutant oligonucleotides were annealed and used as cold competitors (5-fold to 1000-fold molar excess) in gel retardation assays. The percent activity in the shifted DNA-protein complexes was calculated using PhosphorImager quantified data and plotted against fold competition calculated on a molar basis.

Probe preparation

Synthetic oligonucleotides homologous to the -410 to -380 region of the *780* gene promoter were used as probes.

The probes were prepared by either 5' end-labeling with kinase or 3' fill-in end-labeling with Klenow fragment. A typical Klenow labeling reaction included 100 ng of DNA, 1 mM each of dCTP, dGTP, and dTTP (except the labeled nucleotide), 100 mCi of α-^{32}P dATP (3000 to 6000 Ci/mmol), or appropriate nucleotide, and 1-5 units (U) of Klenow fragment incubated in Klenow buffer (50 mM Tris-HCl, pH 7.5; 10 mM $MgCl_2$; 1 mM DTT; 5 mg/ml bovine serum albumin, BSA) for 30 min at 37 °C. A typical kinase labeling reaction included 100 ng of DNA, γ-^{32}P ATP (7000 Ci/mmol, 166 mCi), incubated in Kinase buffer (50 mM Tris-HCl, pH 7.6; 10 mM $MgCl_2$; 5 mM DTT; 0.1 mM spermidine; 0.1 mM EDTA) for 1 hr at 37 °C. The oligonucleotide probes were used in gel retardation assays, methylation interference assays, thymine specific interference assays, and in kinetic competition studies.

Methylation interference assay

Methylation of guanines in the binding site of a DNA binding protein often interferes with binding of that protein. Thus, a DNA probe that has been methylated at a position which interferes with binding will not be retarded in a gel retardation assay. This assay was exploited in an effort to identify the specific guanine residues involved in protein-DNA interactions.

A single end-labeled probe was methylated at an average of one site per molecule of probe by the "G" chemical reaction modified from Maxam and Gilbert (Maxam, 1980).

Approximately 1 ng (10^6 counts per minutes (cpm)) of end-labeled probe was dissolved in 5 to 10 ml of Tris-EDTA (TE) buffer (10 mM Tris-HCl, pH 8.0; 1 mM EDTA, pH 8.0). To this probe, 195 ml of DMS reaction buffer (50 mM sodium cacodylate, pH 8.0; 1 mM EDTA) and 4-8 µg of non-specific DNA, mutant probe, or tRNA were added. Five microliters of 10% dimethyl sulfate (DMS) were added, the reaction mixture was vortexed, and then incubated 5 min at 20 °C. At the end of this incubation period the reaction was stopped with DMS stop solution (1.5 M sodium acetate, pH 7.0; 1 M β-mercaptoethanol). The DNA therein was precipitated two times with ethanol, washed with 70% ethanol and dried in a Speed-vac vacuum drier.

The DNA was reconstituted in binding buffer, nuclear extract added and the binding reaction allowed to proceed for 25 min. This reaction was loaded onto a 8 % native polyacrylamide gel and electrophoresed as indicated above. The gel was exposed to autoradiographic film for 5-12 hrs at 4 °C. The protein-DNA complex and the free probe bands were excised from the gel, placed in elution buffer and eluted overnight while shaking at 37 °C. The gel slice was removed from the buffer and the sample then phenol extracted, ethanol precipitated, washed with 70% ethanol, and dried. The DNA probe was suspended in 100 ml of 10% piperidine and incubated at 90 °C for 30 min. This sample was frozen with liquid nitrogen and lyophilized (approximately 2 hrs). The sample was resuspended in 50 ml of water, frozen and lyophilized two

more times (approximately 1 hr each). Cerenkov counts were measured and cpm determined. Formamide loading buffer (Sambrook *et al.*, 1989) was added to the pellet. The sample was heated for 5 min at 95 °C, quick chilled on ice and loaded onto a 10% sequencing gel (Sambrook *et al.*, 1989). The fragments were analyzed for depleted or less intense bands in the complex lane.

Thymine specific modification by $KMnO_4$

Thymine specific DNA modification was performed as described by Rubin and Schmid (1980) with minor modifications. $KMnO_4$ oxidizes the 5-6 double bond in pyrimidines to carboxylic acid and/or aldehyde products resulting in ring opening. Precipitated probe (10^6 cpm) was resuspended in 5 ml of 30 mM Tris-HCl, pH 8.0, denatured at 95 °C for 2 min and then cooled in an ice bucket. Twenty microliters of 0.25 mM potassium permanganate were added and the samples were incubated 10 min at 20 °C. The reaction was stopped with 50 ml of stop buffer (1.5 M sodium acetate, pH 7.0; 1 M β-mercaptoethanol), and 175 ml of nanopure water was added. After two rounds of ethanol precipitation the DNA was dried and resuspended in 10 µl of hybridization buffer (10 mM Tris-HCl, 1 mM EDTA, 30mM NaCl, pH 8.0), incubated at 90 °C for 2-3 min, and slowly cooled to room temperature.

The $KMnO_4$ treated sample was used in a binding reaction and treated as described for the methylation interference

assay. The fragments were analyzed for depleted or less intense bands.

DNase I footprinting protection assay

Cloned wild type oligonucleotide was digested with *Eco*RI, labeled with α-^{32}P dATP as described above, digested with *Sal*I and recovered from a 4% native polyacrylamide gel. This DNA fragment (10^5 cpm) was used in a binding reaction also described above. The DNA-protein complex was separated on an 8% polyacrylamide gel and exposed to autoradiographic film overnight at 4 °C. The bands representing the complex and free probe were excised and incubated in 100 µl of buffer II containing 1 mM $CaCl_2$ and appropriate concentrations of DNase I (0.001, 0.01, 0.1, 1, 2 µg/µl). At the end of a 20 min incubation period, DNase I stop solution was added to terminate the reaction. The DNA was then eluted overnight in elution buffer by shaking at 37 °C. The buffer containing the DNA was separated from the polyacrylamide, extracted with phenol and chloroform, ethanol precipitated and dried. The DNA was dissolved in 10 µl of formamide loading buffer, denatured at 90 °C for two min and loaded onto a 10% sequencing gel containing 7 M urea.

Results

Two complementary 37 base oligonucleotides containing 31 bases of the *780* activator region and 6 bases of a restriction site were synthesized and used in gel retardation assays. The relationship of this oligonucleotide within the context of the *780* activator can be seen in Fig. 2-2. DNA-protein interaction was demonstrated between this oligonucleotide and factors present in the crude nuclear extract made from cauliflower inflorescences. Binding reactions in which proteinase K (Boehringer Mannheim) was included, demonstrated that this mobility shift was the result of DNA-protein interactions (Fig. 2-3).

A series of oligonucleotides with a variety of mutations were synthesized and used in kinetic competition studies. Two mutations, M142 and M144, disrupt the inverted repeat located within this putative regulatory region (-408 to -397). Mutation M142 (-407 and -406) changed the "TG" to "AC". Mutation M144 changed the "CA" to "GT" (positions -399 and -398). The other mutations (M146, M146A, M133, M155, and M157) for this region are described in Fig. 2-4. These mutant oligonucleotides were used as probes and as cold competitors (5-fold to 1000-fold molar excess) in gel retardation assays. Data from competition studies of each mutant can be seen in Figs. 2-5 to 2-8. The total amount of radioactivity was calculated using the PhosphorImager and the

percent activity in the protein-DNA complex was plotted against fold competition (Fig. 2-9). Results of kinetic competition studies showed that the mutations can be divided into two categories: those that disrupt binding and those that have no effect on binding. When bases within the inverted repeat (mutants M142, M144, M146) were altered, specific binding activity was inhibited. When the four bases immediately 3' to the inverted repeat were mutated (M146A), specific binding was also disrupted. Alteration of the intervening bases (M157), the 5' two bp substitution (M155), and the 3' substitution (M133) did not affect specific binding activity.

Upon close analysis of the wild type sequence, limited homology to the well known plant regulatory elements *ocs* (Ellis *et al.*, 1987), G-Box (Guiliano, 1988), AT-rich, and the light responsive elements (Box II, Box III) (Green *et al.*, 1987) is observed. The *ocs*-element has been identified in the promoter regions of six T-DNA genes and three plant viral genes (Ellis *et al.*, 1987; Bouchez *et al.*, 1989; Fox *et al.*, 1992). OCSBF-1 cloned from a maize root tip cDNA expression library has been shown to bind the *ocs*-element (Singh *et al.*, 1990). A protein factor from pea nuclear extract, GT-1, has been shown to bind light responsive elements found in the promoter of the pea *rbcS-3A* gene at boxes II and III (Green *et al.*, 1987). The G-box element has been shown to interact with numerous plant DNA binding proteins and transcription factors (Williams *et al.*, 1992).

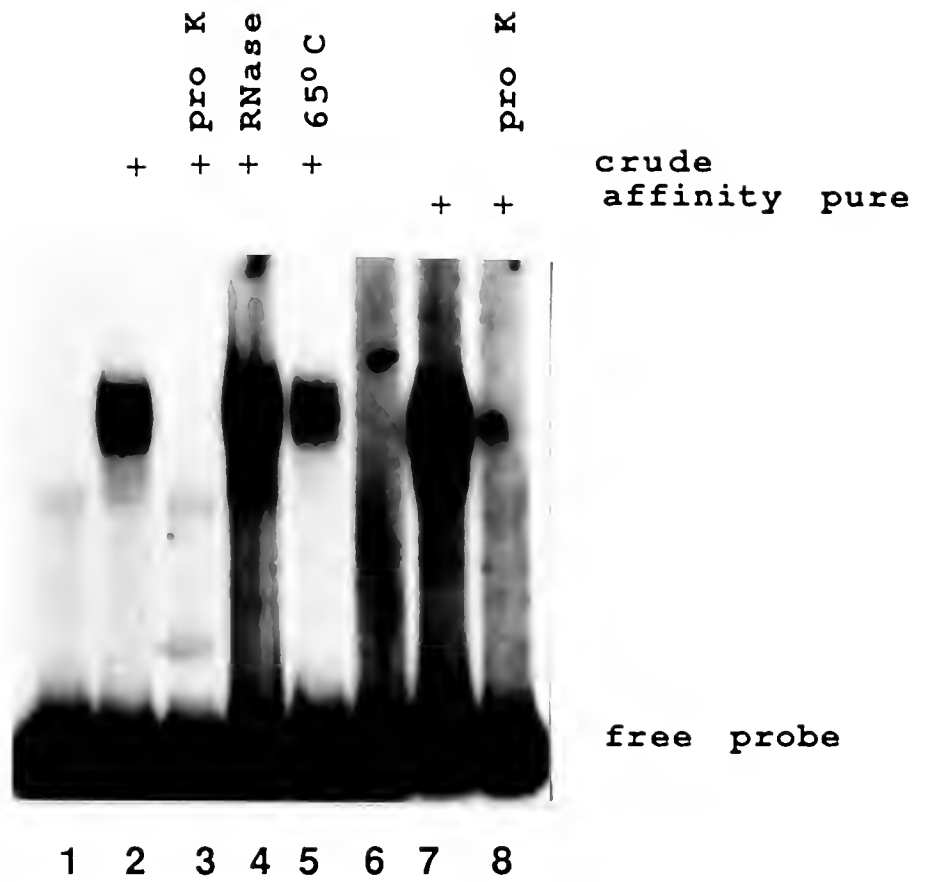


Figure 2-3. Gel retardation assay of wild type oligonucleotide using cauliflower nuclear extract. Approximately 1 ng of wild type probe (1 X 10^5 cpm) was incubated with 1 µl of extract (lanes 2, 3, 4, 5) and with 1 µl of double-affinity pure protein (lanes 7, 8). Lanes 1 and 6 contain probe alone. Lanes 3 and 8 contain proteinase K. Lane 4 contains RNAse and lane 5 contains extract incubated at 65 °C for 10 min.

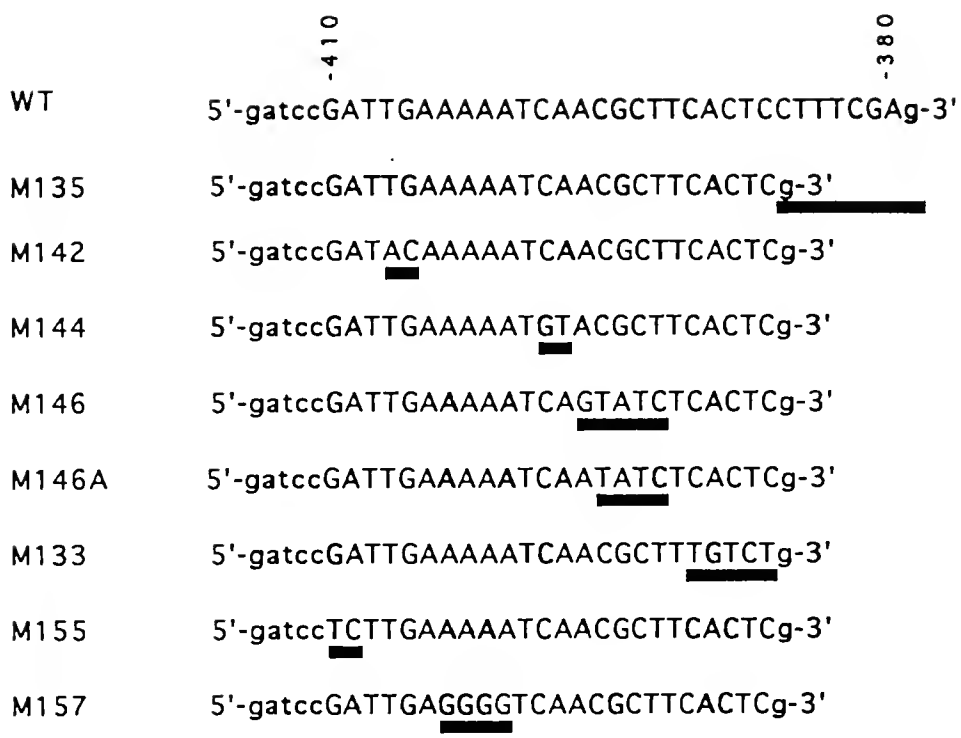


Figure 2-4. Wild type (WT) and mutant oligonucleotide sequences used in competition studies. The bars underscore the mutation in each oligonucleotide. Complementary strands for mutant oligonucleotides were synthesized and annealed by heating equal concentrations of each strand at 90 °C for 3 min and then incubating at 65 °C for 1 hr in annealing buffer.

Figure 2-5. Competition studies with wild type (non-labeled) (A) and M133 (B) competitors. Fold competition was as indicated. Each lane contained 0.15 ng of wild type oligonucleotide probe and 3 µl (approximately 6 ug) of crude nuclear extract. Gel retardation assays were performed as indicated in Materials and Methods.

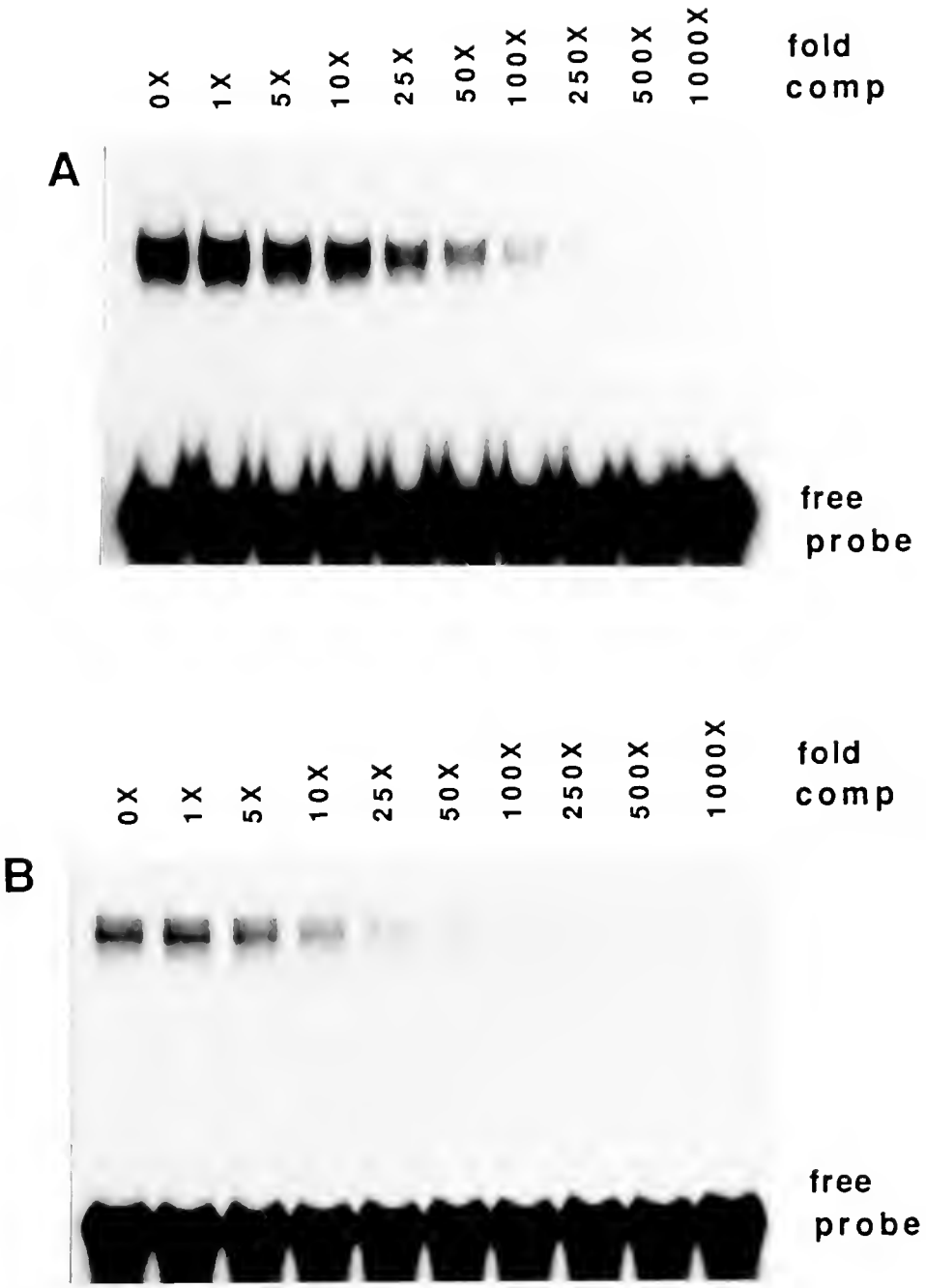
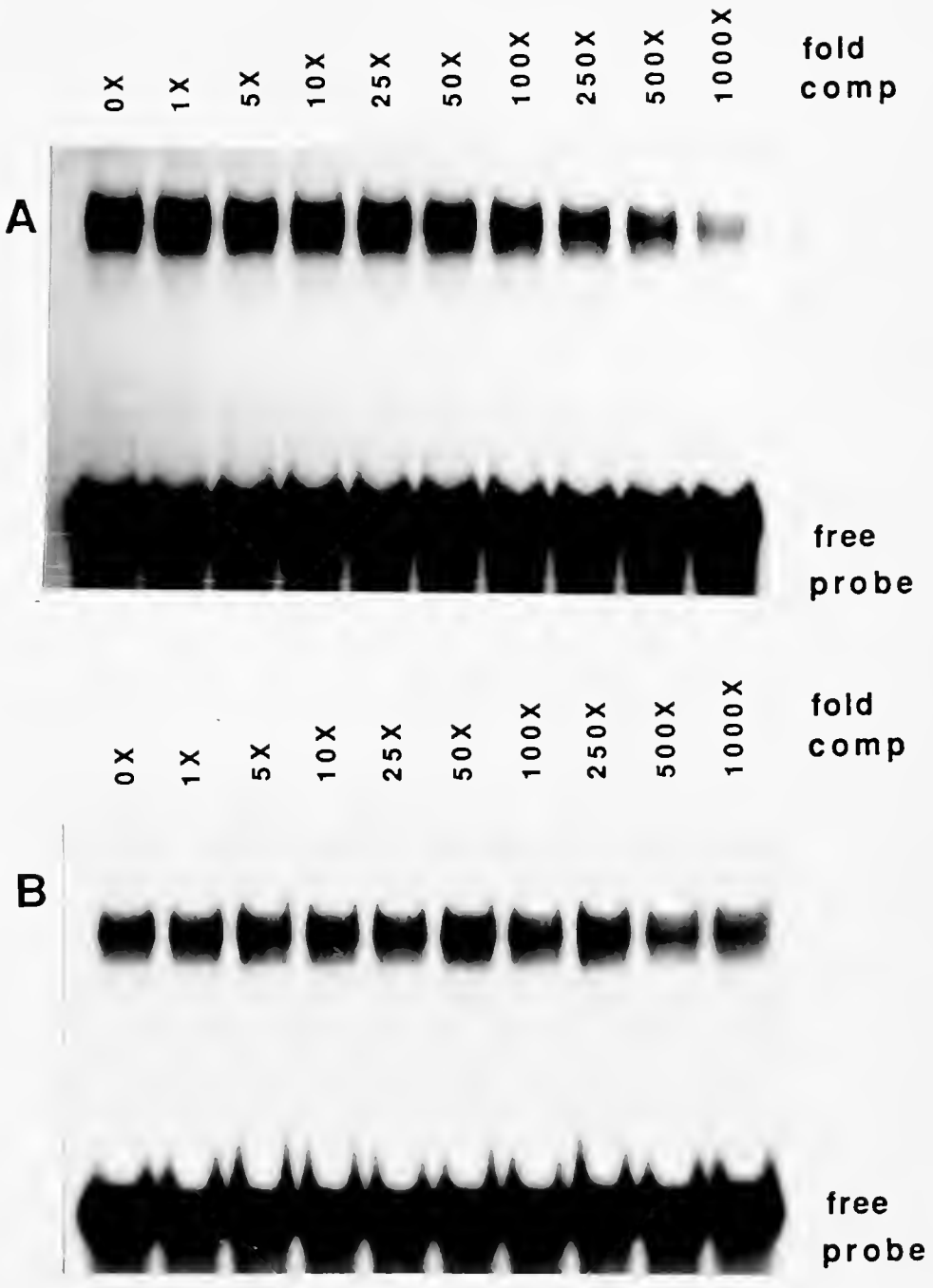

0X
1X
5X
10X
25X
50X
100X
250X
500X
1000X
fold comp
A
free probe
0X
1X
5X
10X
25X
50X
100X
250X
500X
1000X
fold comp
B
free probe

Figure 2-6. Competition studies with M142 (A) and M144 (B) as competitors. Fold competition was as indicated. Each lane contained 0.15 ng of wild type oligonucleotide probe and 3 µl (approximately 6 µg) of crude nuclear extract. Gel retardation assays were performed as indicated in Materials and Methods.


0X
1X
5X
10X
25X
50X
100X
250X
500X
1000X
fold comp
A
free probe
0X
1X
5X
10X
25X
50X
100X
250X
500X
1000X
fold comp
B
free probe

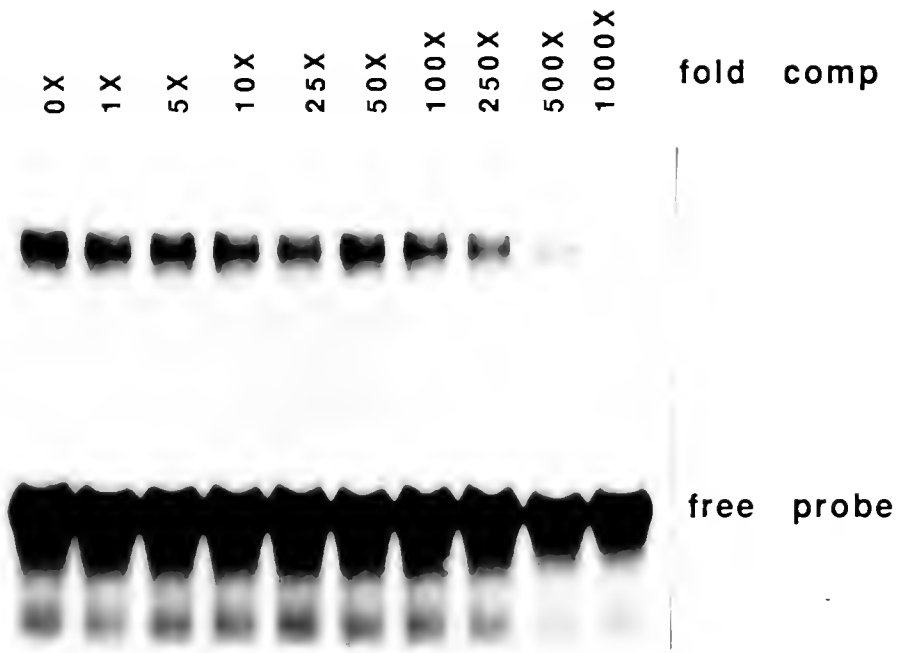


Figure 2-7. Competition studies with M146 as competitor. Fold competition was as indicated. Each lane contained 0.15 ng of wild type oligonucleotide probe and 3 µl (approximately 6 µg) of crude nuclear extract. Gel retardation assays were performed as indicated in Materials and Methods.

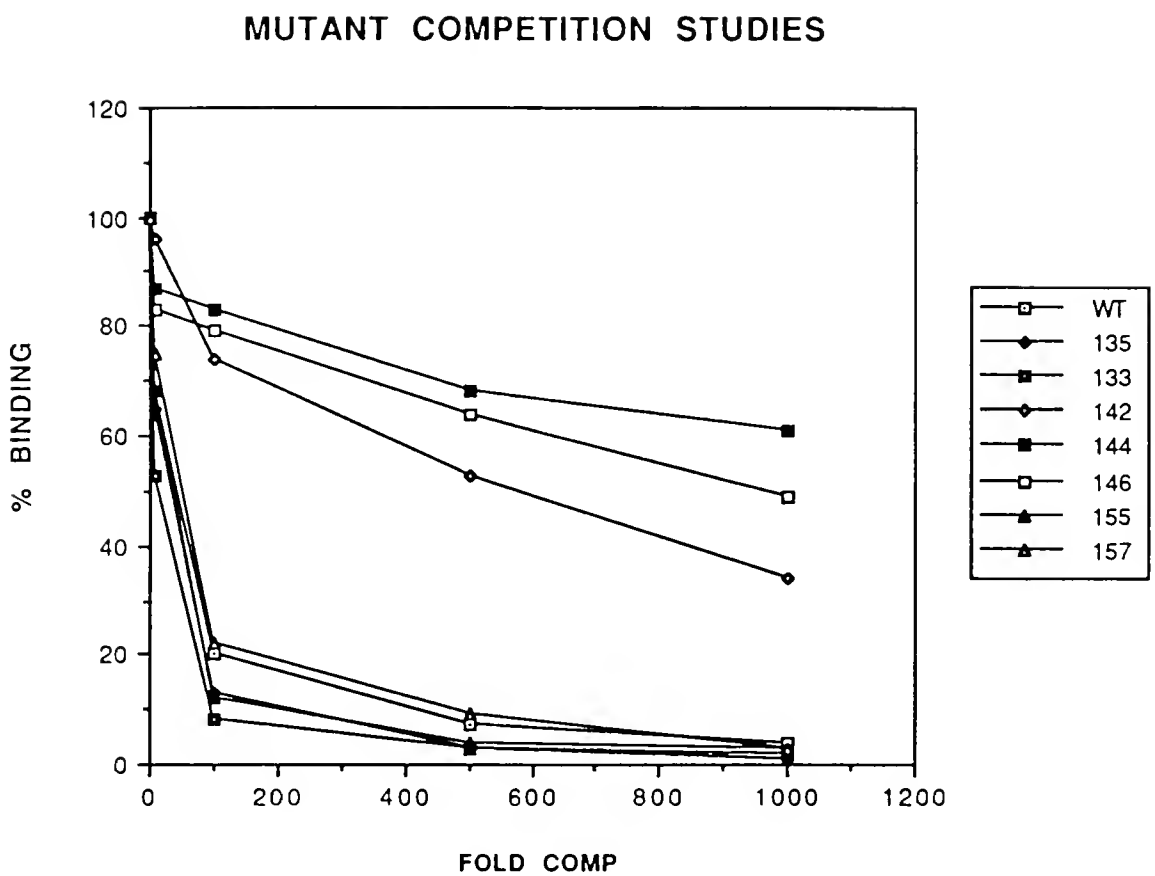


Figure 2-8. Summary of mutant competition studies.

To demonstrate that the specific activity observed in cauliflower inflorescences is indeed a novel activity, a series of oligonucleotides representing these heterologous plant elements were synthesized (Fig. 2-9). An AT composite element (*ATcom*) shown to bind AT factors from a variety of plant nuclear extracts (Czarnecka *et al.*, 1992) was also synthesized. These oligonucleotides were used as competitors in the kinetic competition studies as described above (Fig. 2-10 to 2-12). In order to assess the specificity of binding, the percent activity in the complex was quantitated and plotted against fold competition (Fig. 2-13). The specific 780BP binding activity was not competed off by up to 1000 fold molar excess of each of the heterologous competitors.

An observation made as a result of numerous gel retardation assays was the appearance and disappearance of a doublet banding pattern for the protein-DNA complex. This doublet was not seen when the gel retardation assay was performed with freshly prepared or with once-thawed nuclear extract. The doublet was observed when either the nuclear extract had undergone a number of purification steps at room temperature, or when the extract had undergone numerous rounds of freezing and thawing. This pattern would suggest that the doublet was the result of protein degradation. It is also possible that this doublet resulted from protein-protein interaction between the 780 binding protein (780BP)

HETEROLOGOUS COMPETITORS

C-112	5'-gatcTGACGTAAGCGCTTACGTCA - 3'
G-Box	5'-GCCACGTGGC-3' (4X)
A/Tcom	5'-tcgacAAAAATAATATTAATATTATATTGAAAg-3'
Box II	5'-GTGTGGTTAATATG-3' (4X)
Box III	5'-ACTTTATCATTTTCACTATCT-3' (4X)

Figure 2-9. Heterologous oligonucleotide sequences used in competition studies. Complementary strands for A/Tcom, BoxII, and BoxIII were synthesized and annealed by heating equal amounts of complementary strands at 90 °C for 3 min and then incubating at 65 °C for 1 hr in annealing buffer. The C-112 and the G-box oligonucleotides were self annealed under the same conditions.

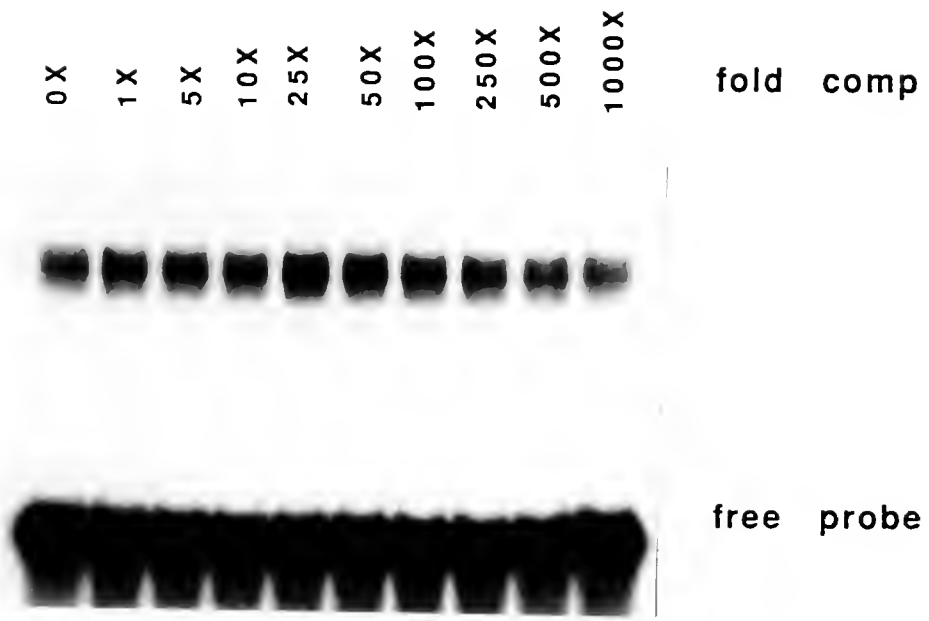


Figure 2-10. Competition studies with A/Tcom. Fold competition was as indicated. Each lane contained 0.15 ng of wild type oligonucleotide probe and 3 µl (approximately 6 µg) of crude nuclear extract. Gel retardation assays were performed as indicated in Materials and Methods.

Figure 2-11. Competition studies with *ocs*-element (A) and G-box (B) as competitors. Fold competition was as indicated. Each lane contained 0.15 ng of wild type oligonucleotide probe and 3 µl (approximately 6 µg) of crude nuclear extract. Gel retardation assays were performed as indicated in Materials and Methods.

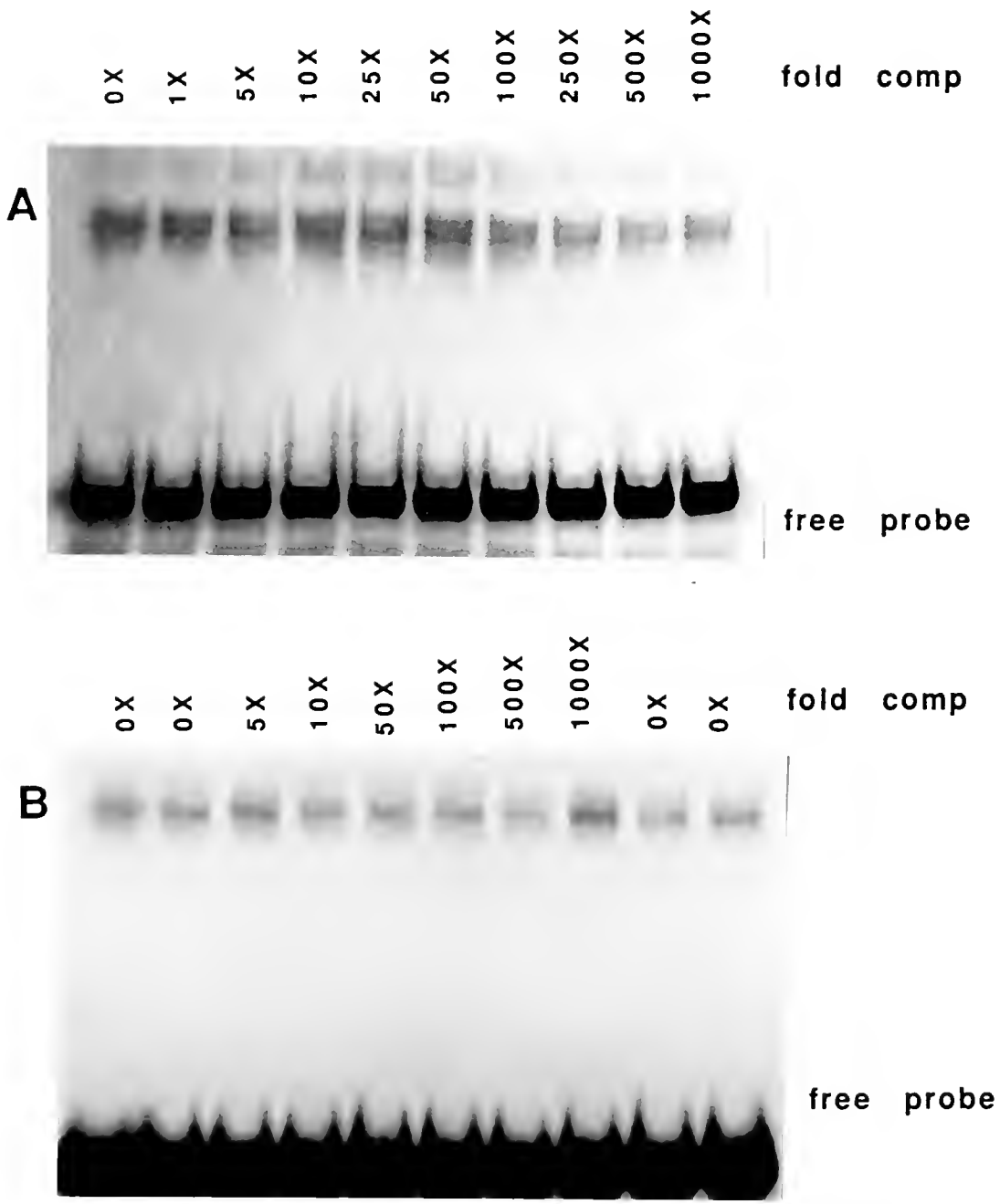
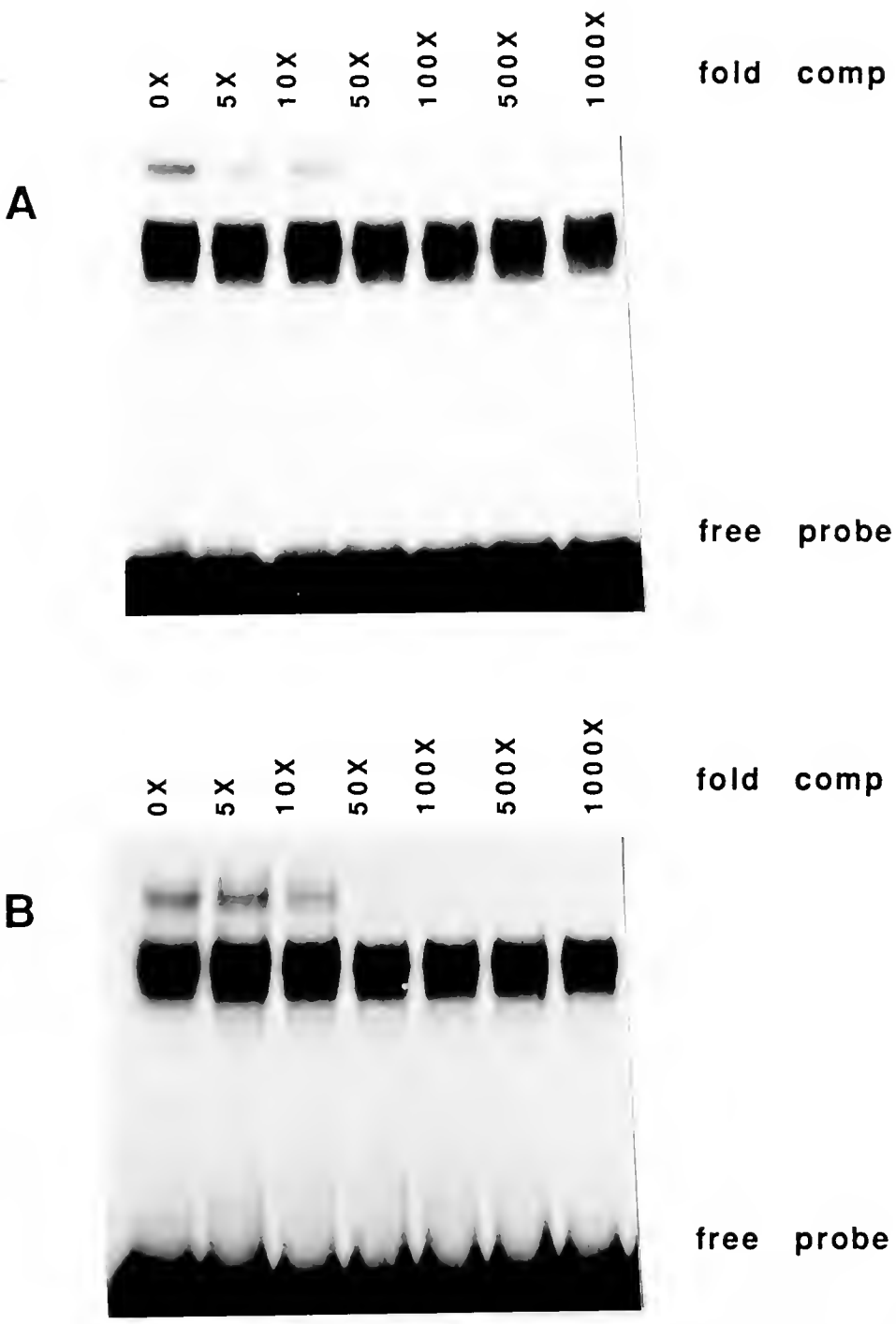
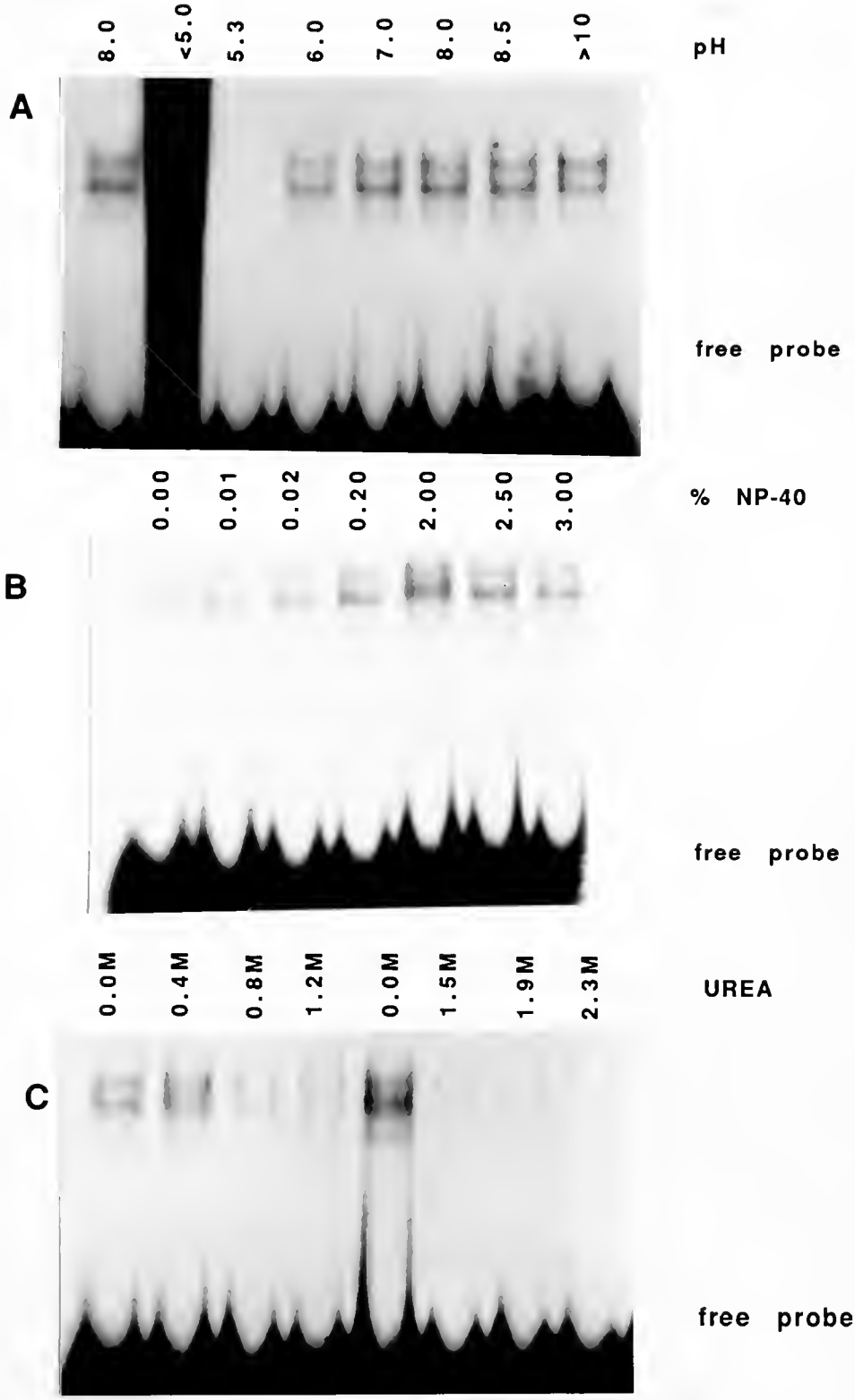

0X
1X
5X
10X
25X
50X
100X
250X
500X
1000X
fold comp
A
free probe
0X
0X
5X
10X
50X
100X
500X
1000X
0X
0X
fold comp
B
free probe

Figure 2-12. Competition studies with Box II (A) and Box III (B) as competitors. Fold competition is as indicated. Each lane contains 0.15 ng of wild type oligonucleotide probe and 3 µl (approximately 6 µg) of crude nuclear extract. Gel retardation assays were performed as indicated in Materials and Methods


A
0X
5X
10X
50X
100X
500X
1000X
fold comp
free probe
B
0X
5X
10X
50X
100X
500X
1000X
fold comp
free probe

and a smaller protein, or from the presence of two distinct proteins with affinity for the probe. If this doublet was the result of protein-protein interaction, then this interaction should be disrupted by biochemical agents known to cause protein conformational change. Mosser and colleagues (1990) demonstrated that incubation of HeLa cytoplasmic extract under biochemical conditions of increased hydrogen ion, urea, and nonionic detergent concentration decreases the temperature dependence for heat shock transcription factor (HSF) activation *in vitro*. Their hypothesis is that HSF possesses DNA binding capability before activation, but is in a complex with another protein, such as HSP70, which blocks the DNA binding domain. Thus, this activation is caused by disruption of protein-protein interactions. The incubation of disruptive agents with cauliflower nuclear extract should have eliminated protein-protein interactions between 780BP and smaller proteins resulting in the disappearance of the doublet in favor of a single band. In the experiment shown in Fig. 2-14, the binding of a second protein to the 780BP can be ruled out as a cause of the double bands since an increase in the concentration of hydrogen ions, urea, or NP-40, in gel retardation assays had no effect on doublet band formation. The same observation was made when Ca^{2+} was used (data not shown). The possibility of two separate DNA binding proteins simultaneously binding the same probe was ruled out in an experiment in which the concentration of extract was varied

Figure 2-14. Gel retardation assays indicating effects of pH (A), NP-40 (B) and urea (C) on doublet stability. Under normal binding conditions the pH is 8 and the NP-40 concentration is 0.01 %.

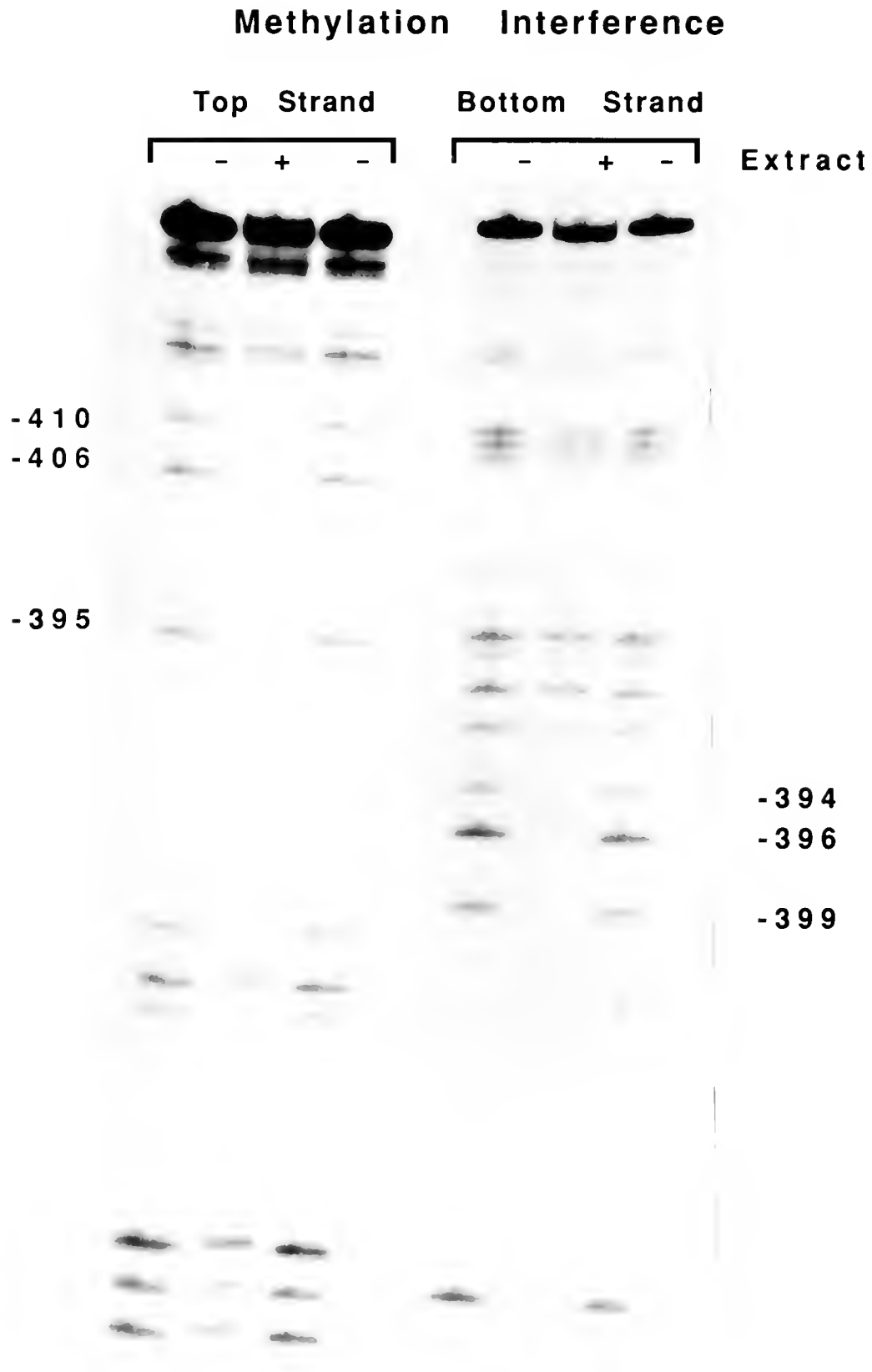

8.0
<5.0
5.3
6.0
7.0
8.0
8.5
>10
pH
A
free probe
0.00
0.01
0.02
0.20
2.00
2.50
3.00
% NP-40
B
free probe
0.0M
0.4M
0.8M
1.2M
0.0M
1.5M
1.9M
2.3M
UREA
C
free probe

and no change in the number of bands was observed. The results of these experiments are consistent with two possible explanations for the doublet pattern: protein degradation of one 780BP, or two 780BPs that show mutually exclusive binding to the probe. However, precautions were taken to decrease the possibility of protein degradation and all purification steps were conducted at 4 °C.

The results of DNase I footprinting (Figs. 2-17 and 2-18), the methylation interference assay (Fig. 2-15) and the $KMnO_4$ interference assay (Fig. 2-16) further define the 780BP binding site. A methylated *Eco*RI/*Sal*I fragment was used to define G residues in close contact with the 780BP. The G residue at positions -410 and -406 on the top strand, and the G residues at -399, at -396, and at -394 on the bottom strand are of major significance for binding of the 780BP. The G residue at -395 on the bottom strand was shown to be of minor significance (Fig. 2-15).

Rubin and Schmid (1980) developed reaction conditions with $KMnO_4$ in which thymine residues in single stranded DNA would be preferentially modified. Thymine specific DNA modification by $KMnO_4$ combined with gel retardation assays was used by Truss and colleagues (1990) to demonstrate that modified T's within the hormone responsive element interfered with binding of the progesterone receptor. This technique was used here to identify specific T residues important for the binding of the 780BP. The T residues at -400 on the top strand and at five positions from -401 to -405 on the bottom

Figure 2-15. Methylation interference assay. Positions of methylated G residues are indicated on both the top and the bottom strands. Methylation interference assays were performed as described in the Materials and Methods. Lanes with (+) and without (-) extract are indicated.


Methylation Interference
Top Strand
Bottom Strand
- + - - + -
Extract
-410
-406
-395
-394
-396
-399

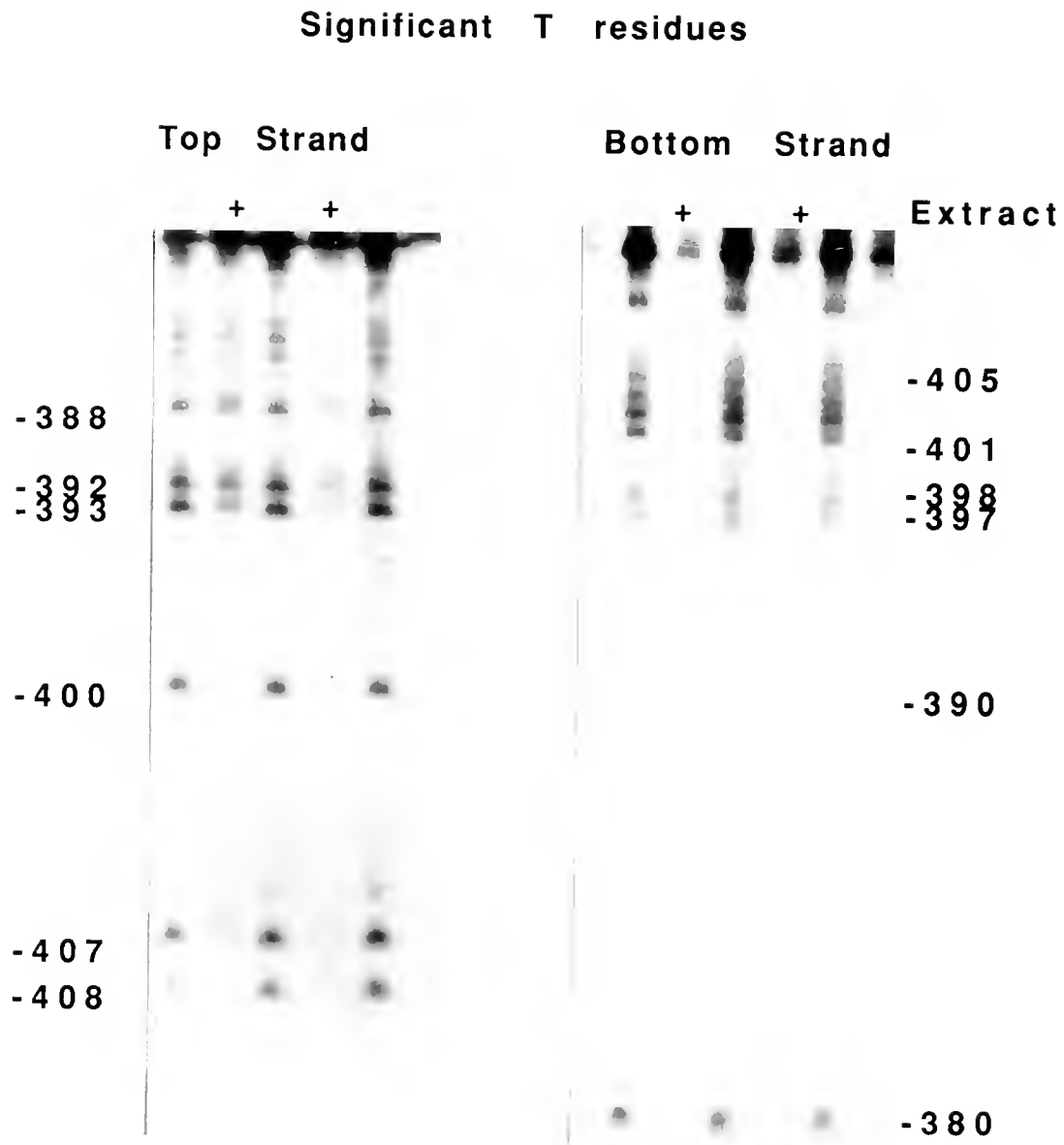


Figure 2-16. $KMnO_4$ interference assay. Significant T residues are indicated.

Figure 2-17. DNase I footprint using the wild type oligonucleotide probe (43 bp). The 5' end of the top strand was labeled with Kinase. This probe (0.3 ng) was incubated with approximately 90 µg of crude nuclear extract, 200 ng of A/Tcom oligonucleotide, 1 µg each of tRNA and poly (dI-dC) in binding reaction. The gel retardation assay was run as described in Materials and Methods for DNase I footprinting, technique I. Lane 1, T ladder; lane 2, free probe; lane 3, free probe , DNase I; Lanes 4, 6, 7, 9, uncomplexed free probe, DNAse I; lanes 5 and 8, probe, extract, DNase I. DNase I concentration in lanes 3-6 was 0.4 ng/µl and in lanes 7-8, 1.6 ng/µl. The protected region is outlined with an open rectangle in the margin.

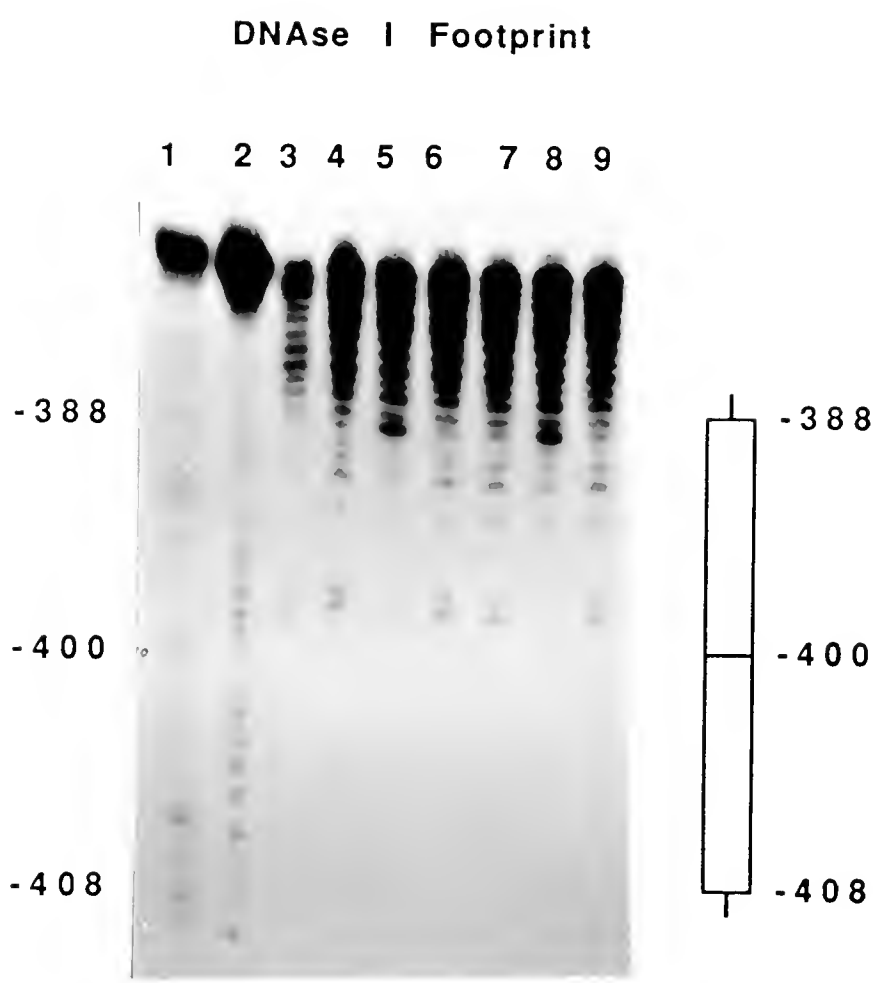
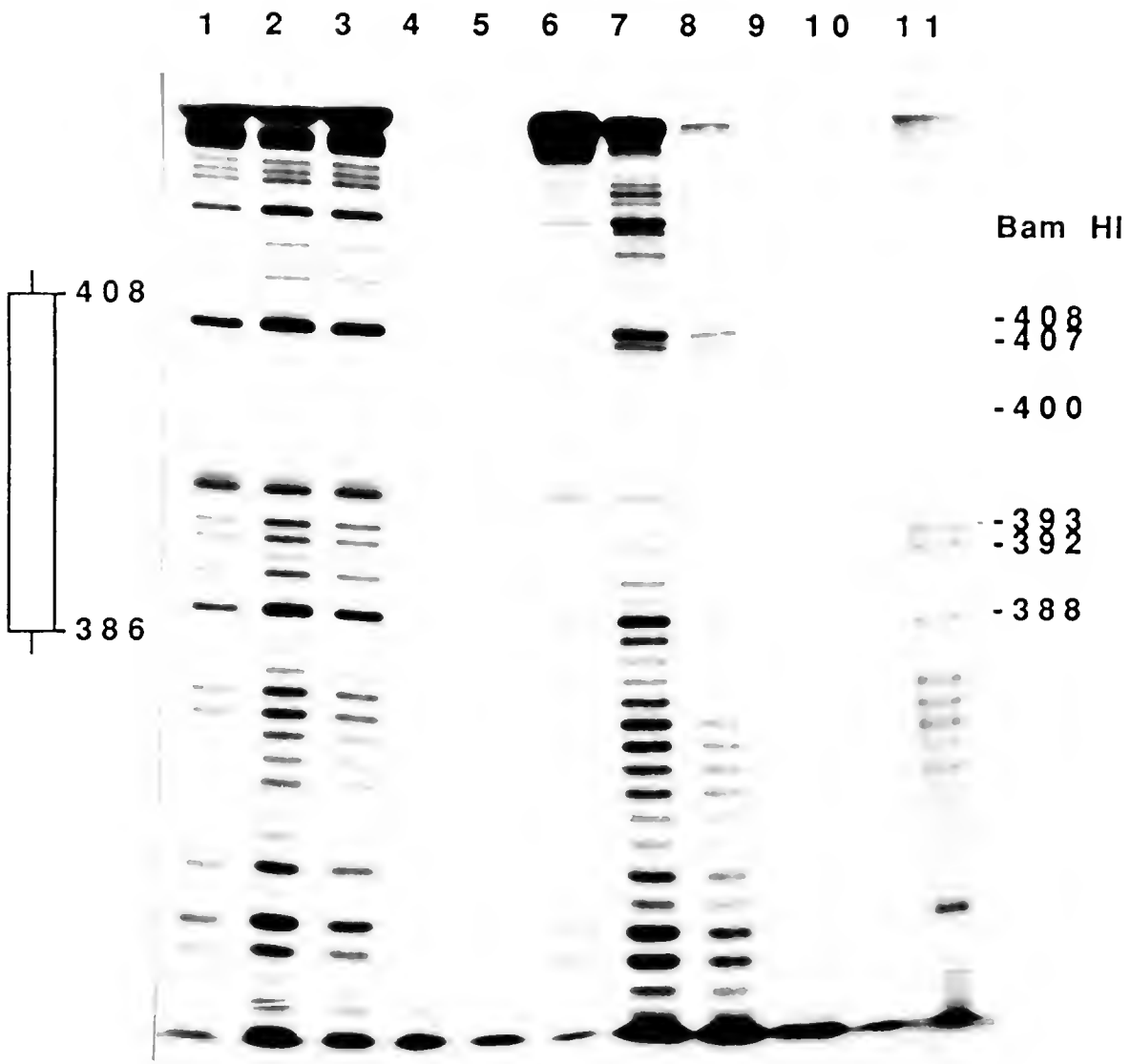

DNAse I Footprint
1 2 3 4 5 6 7 8 9
-388
-400
-408
-388
-400
-408

Figure 2-18. DNase I footprint using the *Eco*RI/*Sal*I probe (72 bp). The probe was labeled with the Klenow fragment by the 3' fill-in procedure described in Materials and Methods. Top strand (2 X 10^5 cpm) probe was incubated with approximately 90 µg of crude nuclear extract, 200 ng of A/Tcom oligonucleotide, 2 µg each of tRNA and poly (dI-dC) in binding reaction. DNase I was added to final concentrations of 0.001, 0.01, 0.1, 1, 2 µg/µl and the reaction continued as described in Materials and Methods, DNase I footprinting technique II. Lanes 1-5 contain free probe and increasing concentrations of DNase I; lanes 6-10 contain extract, probe and increasing concentrations of DNase I. Protected region can be seen when lanes 2 and 7 are compared, and is outlined with a rectangle in the margin.

DNAse I Footprint #2



strand were shown to be very significant for binding (Fig. 2-16). T residues at -392, -393, and -407 on the top strand and at -398 on the bottom strand have minor significance (Fig. 2-16). Since the substitution of 5'-AAAA-3' with 5'-GGGG-3' did not inhibit binding, the apparent interference of the T residues seen here may have been the result of protection of this region by the bound factor and not from specific interaction with the T residues.

DNAse I footprinting is perhaps the most widely used method for the identification of specific protein binding sites on a DNA fragment (Figs. 2-17 and 2-18). The demonstration of a clean DNase footprint proved very difficult for this fragment. In Fig. 2-17, an oligonucleotide probe was used in this analysis. Protection is apparent in lanes 5 and 8 in the region from -399 to -388. A hypersensitive site is observed in the area of -388, indicating the boundary of a site protected by a protein. Distinct protection is visible between -399 and -396. This protection is within the region shown to be necessary for binding by the methylation interference and the $KMnO_4$ interference assays. The absence of DNase cleavage 5' of -400 was due to the short length of the probe and/or to less than optimum incubation time of fragment with DNase I. Fig. 2-18 shows the results of another attempt to DNase I footprint. In this experiment, the oligonucleotide was cloned into the *BamHI* site of the pUC19 polylinker and digested with *EcoRI* and *Sal*I to produce a larger fragment

that was used in this experiment. In comparison of lane 2 (no extract) with lane 7 (extract added), protection can be seen from -388 to -410. And as with footprint number 1, hypersensitivity can be seen near the -388 site.

Discussion

A 16 bp sequence of the *780* activator region has been shown to bind a nuclear protein found in cauliflower extracts. The site of protein binding complete with the identification of significant G and T residues is presented in Fig. 2-19. This sequence, 5'-TTGAAAAATCAACGCT-3', is located between -408 and -393 and is disrupted by the -396 5' deletion described by Bruce and Gurley (1987). The two 5' deletions to positions -427 and -396 resulted in a relative transcription level of 92% and 55%, respectively, when compared to wild type. The sharp drop in promoter activity obtained by deleting sequences between -427 and -396 strongly suggests that a transcriptional regulatory element is located within, or overlaps with, the 31 bp between these two deletions. Because this sequence alone has not been shown directly to be a regulatory element, the binding site for the 780BP located within this region will hereafter be referred to as the "*780* binding element" (780BPE).

The significance of the cluster of the "abc" repeats, also located within this region, has not been determined. The role of these repeats as possible transcriptional regulatory elements was recently addressed by O'Grady (1993).

The "c" repeat is located within this region, defined as the 780BPE, and seems to be a significant component of the 780BPE.

Interpretation of competition studies with a series of 2 bp and 5 bp substitution mutations resulted in the development of a model for DNA-protein interaction involving two half sites. The 5'-AAAA-3' to 5'-GGGG-3' substitution mutation did not result in the disruption of the 780BP binding activity. Mutations M142, M144, and M146 lie on the 5' and 3' side of this substitution mutation and as summarized in Fig. 2-8, each of these mutations disrupted binding of the 780BP. The other mutations, M133, M135, and M155, did not disrupt binding. These mutations, therefore, lie outside of the defined 16 base element both at the 5' and the 3' end.

The fact that specific DNA-protein interactions were not disrupted when a series of heterologous plant transcription elements were used as competitors is evidence that the 780BP activity is a novel activity. The limited homology to the well known plant transcription elements G-box (Guiliano, 1988), *ocs* (Ellis *et al.*, 1987), GT-1 binding box II and box III (Green *et al.*, 1987), and to AT-rich sequences (Jensen *et al.*, 1988; Czarnecka *et al.*, 1990; Jacobsen *et al.*, 1990) proved to be of no consequence.

Upon initial reading, it would seem that conflicting results are being presented here. From the mutation studies (Fig. 2-9), mutant M155 which substitutes TC for GA at

position -410 and -409 did not affect the binding of 780BP. However, the methylation interference assay (Fig. 2-15) indicates that methylation of the G residue at position -410 interfered with the binding of a factor in the crude nuclear extract. Perhaps this latter result can be explained by the closeness of this G residue to the site of specific DNA-protein interaction. This argument can also be used to explain more conflicting results observed in the mutation studies (Fig. 2-9) which showed that the M157 mutant (5'-GGGG-3' substituted for 5'-AAAA-3') did not interfere with the binding of 780BP, and the $KMnO_4$ interference assay indicated that the corresponding T residues on the bottom strand were important for binding.

The observation of a doublet binding activity periodically raised the question of whether one or two proteins were interacting with this binding site. The doublet could have resulted from the instability of protein-protein interactions between two proteins, one with DNA binding capability and the other with no DNA binding capability. This hypothesis was dispelled by the results of experiments which exposed crude nuclear extract to a series of chemical reagents heretofore shown to disrupt protein-protein interactions. The doublet once formed persisted throughout these experiments. When binding was affected by these conditions, both bands were affected uniformly. It was concluded that the doublet was the result of protein degradation or the interaction of two closely related

proteins exhibiting mutually exclusive binding to the same site.

The nature of the sequence, an inverted repeat separated by four bases, causes one to speculate as to the motif that might comprise the DNA binding domain of the 780BP. The bZIP motif which has been identified in most plant transcription factors is not a consideration. Proteins with a bZip DNA binding domain bind sequences comprised of bases that form a "dyad symmetrical binding site" (Vinson *et al.*, 1989). The question of two independent binding sites is dispelled by the results of the mutational analysis.

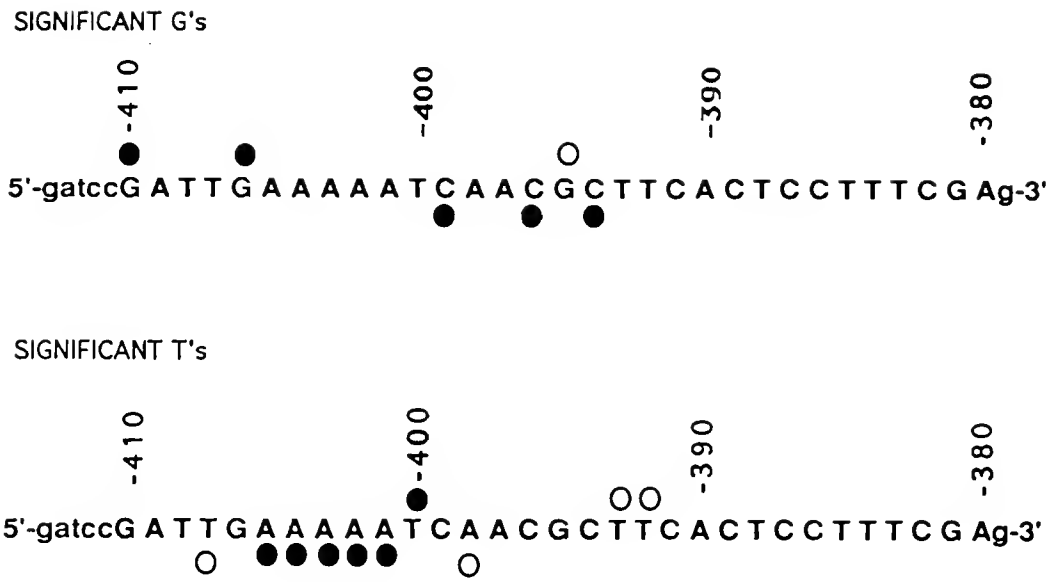


Figure 2-19. 780BPE complete with signicant G and T bases. Shaded circles indicate bases that were strongly protected by 780BP. Open circles indicate bases that were weakly protected.

CHAPTER 3

PURIFICATION OF THE *780* BINDING PROTEIN (780BP)

Introduction

Standard methods of protein purification have been used frequently for the separation of animal transcription factors from each other and from other nuclear proteins. The promoter selective transcription factor Sp1 was purified from human cells to more than 95% homogeneity by sequential column chromatography and by sequence-specific DNA affinity chromatography (Briggs *et al.*, 1986; Kadonaga & Tjian, 1986). The DNA binding activity was monitored throughout the purification process by DNase I footprinting experiments. The first chromatographic step was Sephacryl S-300 gel filtration followed by DEAE Sepharose, heparin agarose, fast performance liquid chromatography (FPLC) Mono S, and DNA affinity. The specific activity increased from 17 units of DNA binding activity per mg (U/mg) in nuclear extract to 1,200 U/mg in FPLC Mono S fractions. The specific activity of the DNA affinity fractions increased approximately 90 fold over that of the Mono S fractions (1,200 U/mg to 100,000 U/mg) dramatically demonstrating the power of affinity chromatography.

Three proteins which bind to different sites on the mouse immunoglobulin heavy-chain enhancer were chromatographically separated and partially purified from mouse plasmacytoma cell line MPC11 (Peterson & Calame, 1987).

Specific activity was monitored throughout the purification process by the gel retardation assays. The binding sites were precisely mapped by methylation interference, DNase I footprinting, and orthophenanthroline/copper (OP/Cu) chemical nuclease footprinting. DNA affinity chromatography was used in a later report to purify the immunoglobulin heavy chain enhancer site C2 (mE3) binding protein to homogeneity (Peterson & Calame, 1989).

Drosophila heat shock transcription factor (HSF) was purified to 95% homogeneity by sequence specific affinity chromatography from nuclear extract prepared from heat shocked cells (Wu *et al.*, 1987). Nuclear extract was chromatographed by heparin-Sepharose, twice by DNA affinity, and FPLC Mono S chromatography. Photo-affinity labeling (UV-crosslinking) (Lin & Riggs, 1974) combined with SDS gel electrophoresis was used to determine a molecular weight of 110 kD for this heat HSF. UV-crosslinking was also used to measure the molecular weight of the adenovirus major late transcription factor (Chodosh, 1986).

Reports on the purification of plant DNA-binding proteins and transcription factors are scarce. Two reports, one on the partial purification of a G-box binding protein from *Arabidopsis* suspension culture (DeLisle & Ferl, 1990), and the other on the purification of the silencer binding factor (SBF-1) from bean suspension culture (Harrison, 1991) have been published. Sequence specific SBF-1 activity was purified approximately 1750-fold from suspension cell nuclei

using a series of ammonium sulfate precipitations, gel filtration (Sephacryl 300), heparin-agarose and sequence-specific DNA affinity chromatography (Harrison, 1991). Using a tetramer of the sequence specific binding site, specific binding activity was monitored by gel-retardation analysis of fractions during the purification process. A molecular weight of 160-200 kD was determined for this protein in solution. A subunit molecular weight of 95 kD was determined by SDS polyacrylamide gel electrophoresis and UV-crosslinking. The apparent discrepancy in molecular weight suggests that SBF-1 exists as a dimer in solution and may possibly bind as a dimer. The instability of the affinity purified factor prohibited DNase I footprinting. However, DNase I protection was demonstrated with heparin-agarose partially purified protein.

While the expressed purpose of purifying this protein was for microsequencing and antibody production, it is reported in the discussion that the very low abundance and the lack of stability of this factor make it impossible to obtain enough for sequencing. It is reported, however, that polyclonal serum containing antibodies to SBF-1 has been produced. These antibodies could be used for screening cDNA libraries and for future characterizations of this protein activity. To date, the cloning of SBF-1 using the antibody approach has not been reported.

An *Arabidopsis Adh* G-box binding factor was partially purified from whole cell extracts, and from suspension

cultures and leaves by heparin-agarose, gel filtration (Superose 6, Pharmacia), and Mono-Q chromatography (DeLisle & Ferl, 1990). Specific activity was monitored with the gel retardation assay. DNase I footprint analysis was conducted on both crude extract and on heparin agarose fractions. The partial purification of this binding activity seems to be simply directed at the characterization of this protein at various stages of the purification process. Though not stated in the publication, cloning of this factor would certainly seem to be the direction in which these authors were headed.

A series of column chromatography steps has been combined with sequence-specific DNA affinity chromatography (Kadonaga & Tjian, 1986) to purify the 780BP binding activity from cauliflower inflorescence nuclear extract. Among them are heparin Sepharose affinity, and Mono-Q and Mono-S ion exchange chromatography. Heparin Sepharose is used to separate proteins with affinity for negative ligands such as DNA (DNA binding proteins) from those proteins with no affinity for negative ligands. The Mono-Q column is a strong anion exchanger. It is composed of MonoBeads-monodispersed hydrophilic polymer particles for fast, high resolution chromatography. Mono-Q binds negatively charged groups through quaternary amine groups. The Mono-S column is a strong cation exchanger on MonoBeads that bind positively charged biomolecules through sulphonic acid groups. Other techniques such as gel retardation (Fried & Crothers, 1981),

methylation interference (Maxam, 1980; Gilman *et al.*, 1986), DNase I footprinting (Galas & Schmitz, 1978), UV-crosslinking (Lin & Riggs, 1974; Wu *et al.*, 1987), and SDS polyacrylamide gel electrophoresis (Laemmli, 1970) combined with silver staining have been employed to characterize the 780BP.

Materials and Methods

Heparin-Sepharose affinity chromatography

Crude nuclear extract was ammonium sulfate precipitated at 40% saturation. The precipitated protein was resuspended in 0.1 M buffer II and fractionated on a HiTrap Heparin column (Pharmacia) over a 0.1-1 M KCl gradient in buffer II, pH 8.0. All fractions containing the appropriate activity, as indicated by gel retardation assay, were pooled and dialyzed against 0.1 M KCl buffer II. Protein concentrations were determined (Bradford Protein Assay Kit, BioRad) and stored in preparation for the next round of purification. The heparin-Sepharose purification step and all subsequent steps were performed in the cold room (4°C).

Ion-exchange chromatography

Heparin pure nuclear extract was fractionated using a Mono-Q column (Pharmacia) with a 0.1-1 M KCl gradient in buffer II complete with protease inhibitors. All fractions were assayed for specific binding activity. Fractions with specific binding activity were pooled and a small amount

aliquoted and stored for further study. The remainder was loaded onto a Mono-S column (Pharmacia) and was resolved over a 0.1-1M KCl gradient in buffer II. Samples were aliquoted for assay of protein concentration, DNA binding unit determination and molecular weight determination with SDS-PAGE.

Gel filtration chromatography

Gel filtration utilized a Superose 12 column from Pharmacia. Superose 12 is a cross-linked agarose matrix recommended for the resolution of proteins in the 10 kD to the 300 kD range. A 50 ml protein sample was loaded on an equilibrated (buffer II, 200 mM KCl) Superose 12 column and eluted in 25 ml fractions. All fractions were assayed for binding activity. Once the fractions with activity were identified, the relative molecular weight range was determined by comparing the elution profile to that of three molecular weight standards (albumin, 66 kD; carbonic anhydrase, 29 kD; cytochrome c, 12.4 kD; Sigma)

Preparation of DNA affinity resin

A 30 bp double-stranded oligonucleotide containing wild type sequences was used to prepare a DNA-Sepharose affinity resin according to procedure modified from Kadonaga *et al.* (1991). Gel purified, double stranded, wild type synthetic oligonucleotide (250 mg) was phosphorylated with T4 polynucleotide kinase in a reaction containing 3 mM ATP, (pH

7.0) and 5 μCi of γ-^{32}P dATP (used as tracer to monitor coupling to Sepharose) in Kinase buffer (50 mM Tris-HCl, pH 7.6; 10 mM $MgCl_2$; 5 mM DTT; 0.1 mM spermidine; 0.1 mM EDTA). The oligonucleotide was then self ligated in a reaction mixture containing 3 mM ATP (pH 7.0), 30 Weiss units of T4 ligase, 66 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 15 mM DTT, and 1 mM spermidine, and incubated overnight at room temperature.

Coupling of the DNA to cyanogen bromide (CNBr)-Sepharose CL-4B (Pharmacia) was performed as follows. The resin was washed four times with 100 ml of ice-cold water (4 °C), followed by two washes with ice-cold 10 mM potassium phosphate, pH 8.0. Approximately 5 ml of activated resin was transferred to a 15 ml polypropylene screw cap tube to which 2 ml of 10 mM potassium phosphate buffer (pH 8.0), was added. The ligated DNA in 50 ml of water was added and the reaction incubated on a rotating wheel overnight at room temperature. The resin was then transferred to a 60 ml coarse sintered glass funnel and washed two times with 100 ml of water. The final wash sequence was 100 ml of 10 mM potassium phosphate buffer, pH 8.0; 100 ml of 1 M potassium phosphate buffer, pH 8.0; 100 ml of 1 M KCl; 100 ml of water, and finally with 100 ml of column storage buffer (10 mM Tris-HCl, pH 7.6; 1 mM EDTA; 0.3 M NaCl; 0.04% (w/v) Na azide. The resin is reported to be stable for at least a year when stored at 4 °C.

DNA Affinity Chromatography

A uniformly mixed suspension of DNA affinity resin (1 ml) was added to a 1.5 ml Eppendorf tube and equilibrated with 0.1 M KCl in binding buffer (10 mM HEPES, pH 7.9; 10% glycerol; 0.1 mM EGTA pH 7.9; 1.5 mM $MgCl_2$; 3.0 mM DTT; 0.01% NP-40). A 30 sec spin in a table top swinging bucket centrifuge resulted in a packed resin bed volume of 100 µl. To this 100 µl bed volume resin 200 µl of partially purified nuclear extract in 0.1 M KCl binding buffer was added. Binding was allowed to proceed for 15 min while the tube was rotating. The affinity resin was pelleted with a 30 sec spin using a medical table top swinging bucket centrifuge in each step. The supernatant was removed and labeled "bindate". The resin was then washed with 200 ml of binding buffer (100 mM KCl). The supernatant was removed and labeled "100 mM wash". Elution of protein was performed with 400 mM, 700 mM, and with 1 M KCl in the binding buffer. Fractions were labeled appropriately, frozen with liquid nitrogen and stored at -80 °C. Protein concentrations were later determined using the Bradford assay.

Photoaffinity crosslinking

The UV-crosslinking probe (wild type oligonucleotide) was prepared by primer extension. The upper strand was annealed to a 15 base complementary primer and extended with

the Klenow fragment of DNA polymerase with the incorporation of bromodeoxyuridine (BrdU), α-^{32}P dATP, dCTP, and dGTP. The primer extension labeling reaction contained 0.25 mM BrdU:dTTP (1:1), α-^{32}P dATP 3000 Ci/mmol (200 µCi), 0.25 mM dGTP, 0.25 mM dCTP, and 6 U/µl of Klenow fragment. The resulting ^{32}P-labeled probe was used in binding reactions (5 x 10^4 cpm/reaction) with nuclear extract as indicated in the results.

At the end of a 30 min incubation period, the binding reaction was exposed in an open Eppendorf tube to a 302 nM UV transilluminator at a distance of 5 cm (sample to lamp) for 20 min at room temperature. The sample was then prepared and analyzed by SDS-PAGE (Laemmli, 1970). The gel was dried under vacuum and exposed as described above.

Half-life determination

The protein-DNA binding reactions described above was scaled-up 10-fold for half-life determination. After a 30 min incubation of the binding reaction, 100 to 1000 fold of cold competitor DNA was added and mixed. Aliqouts (25 µl) were removed at various time intervals and loaded on a running 8% polyacrylamide gel. The percent labeled probe in the shifted complex was quantified as described previously and plotted against time points as described.

Results

The purification scheme presented here was developed using a small amount of crude nuclear extract from cauliflower florets. The data presented were derived from results obtained from both the small scale and the large scale preparations.

Crude nuclear extract (80-90 mg) was ammonium sulfate precipitated (70% saturation), resuspended in buffer II (0 M KCl), and dialyzed against buffer II containing 100 mM KCl. The extract was then divided into two samples and each sample was separately applied to a Hi-Trap heparin Sepharose column and resolved by stepwise elution (5 ml fractions) with buffer II containing 0.3 M, 0.6 M, 0.8 M, and 1 M KCl, respectively. The absorbance (A_{280}) was monitored for protein elution and fractions were assayed for 780BP activity by gel retardation assays as shown in Fig. 3-1. The 780BP activity eluted between 0.4 M and 0.6 M KCl. The 0.6 M eluate was collected in two 5 ml fractions, numbers 5 and numbers 6. Fraction 6 represented a little less than 50% of the total protein and contained 95% of the 780BP binding activity. This fraction was used in the remaining purification steps.

Fraction 6 from four separate heparin runs were pooled, ammonium sulfate precipitated (70% saturation), dialyzed and applied to a Mono-Q anion exchange column equilibrated with buffer II (0.1 M KCl). Mono-Q flow through fractions with activity (Fig. 3-2) were pooled and applied to the Mono-S

column. DNA binding activity was puried by elution with a 15 ml gradient from 0.1 M to 1 M KCl. Specific binding activity was eluted in three 1 ml fractions at a salt concentration of 0.28-0.46 M KCl (Fig. 3-3).

To further purify this binding activity, a DNA affinity resin was prepared as described in Materials and Methods. Two different techniques were employed with the affinity resin. One technique was the batch method reported above and the second treatment was a column approach. The batch approach proved the most efficient and thus is reported here. The dialyzed Mono-S fraction, which represented 3% of the total protein, was applied to the DNA affinity resin equilibrated with buffer II (0.1 M KCl). It was shown in previous experiments that specific binding activity was completely retained on the affinity resin and eluted between 0.4 M and 0.8 M KCl. To ensure complete recovery of binding activity, elution from the DNA affinity resin was accomplished with 1 M KCl (Fig. 3-4). The presence of non-specific bands in the assay for 780BP activity indicated a need for a second affinity elution. However, before the second affinity experiment was performed, the first affinity pure sample was passed over a Superose 12 gel filtration column. Assay of gel filtration fractions showed specific binding activity to be present in the 13-15 ml fractions (Fig. 3-5). Elution in these fractions corresponds to a molecular weight range of 40-54 kD. The gel filtration fractions were pooled, dialyzed, and subjected to affinity

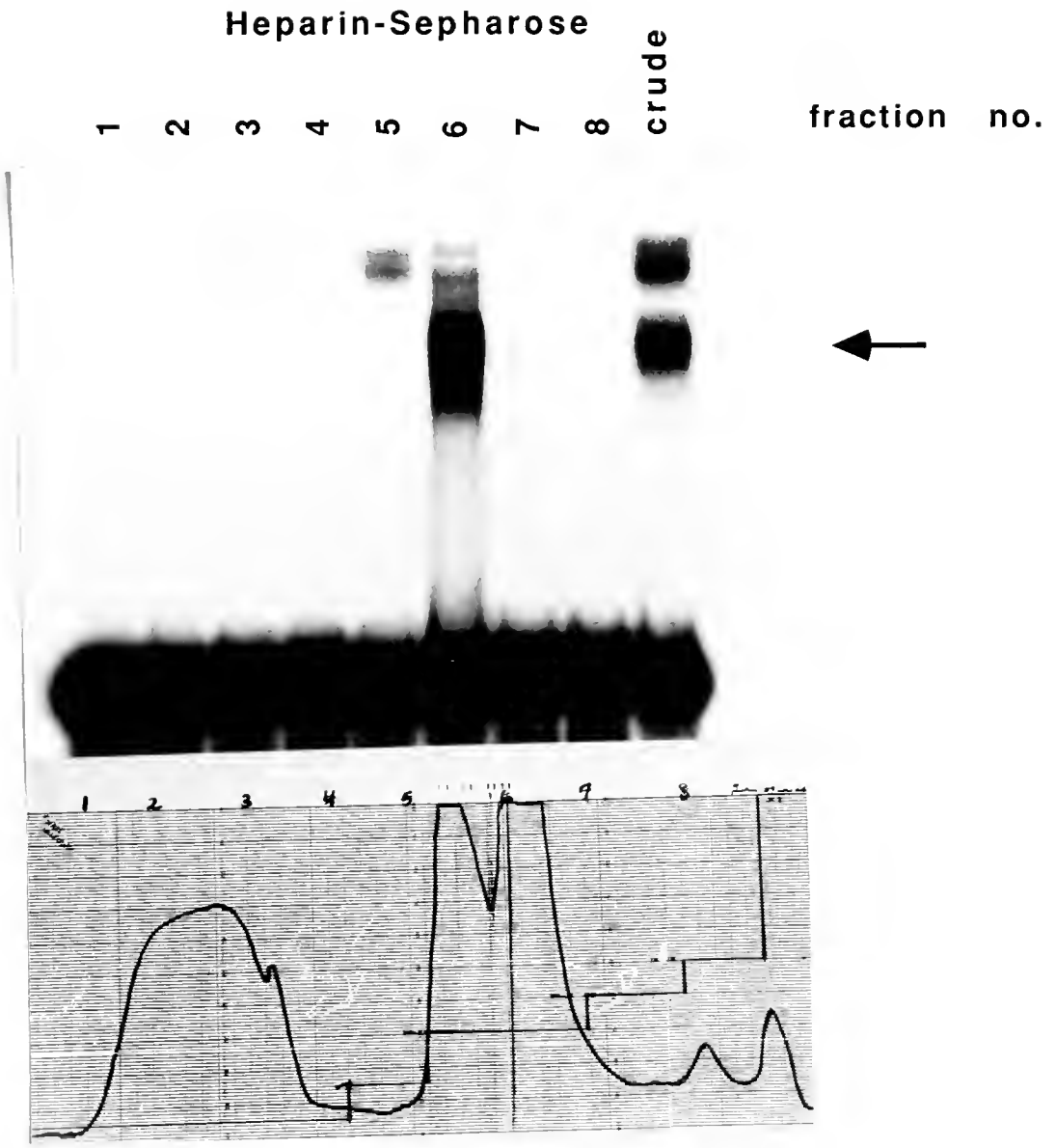


Figure 3-1. Heparin-Sepharose profile. Specific 780BP activity is eluted in fraction # 6. The arrow indicates 780BP activity. The chart at the bottom is a chromatogram showing optical density (A_{280}) peaks from a typical heparin-Sepharose separation.

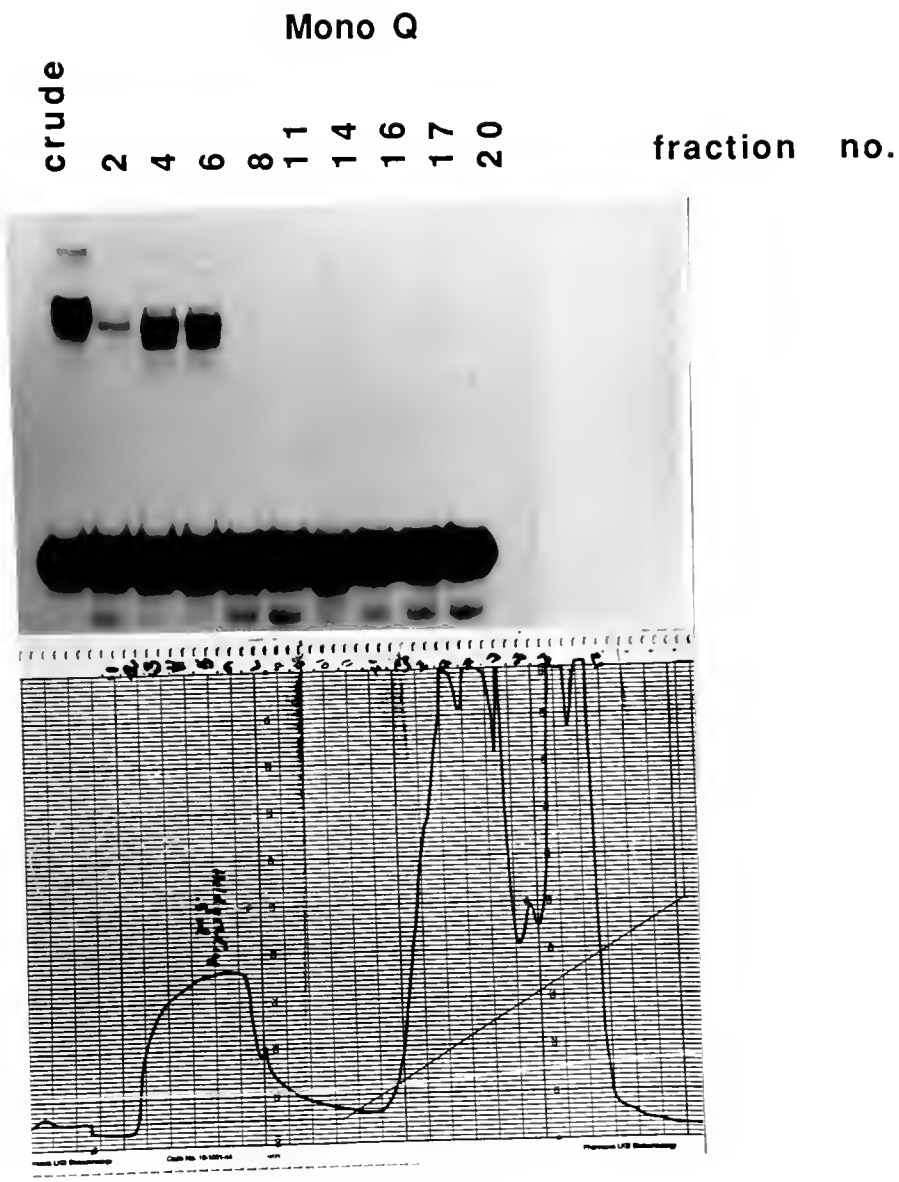


Figure 3-2. Mono-Q elution profile. 780BP activity did not bind Mono Q under these conditions. 1 µl of 1 ml fractions was added to binding reaction and gel retardation assay run as described in Materials and Methods. Specific binding activity was in flow through fractions 2, 4, and 6. The chart below is a chromatogram from a typical Mono-Q separation (A_{280}).

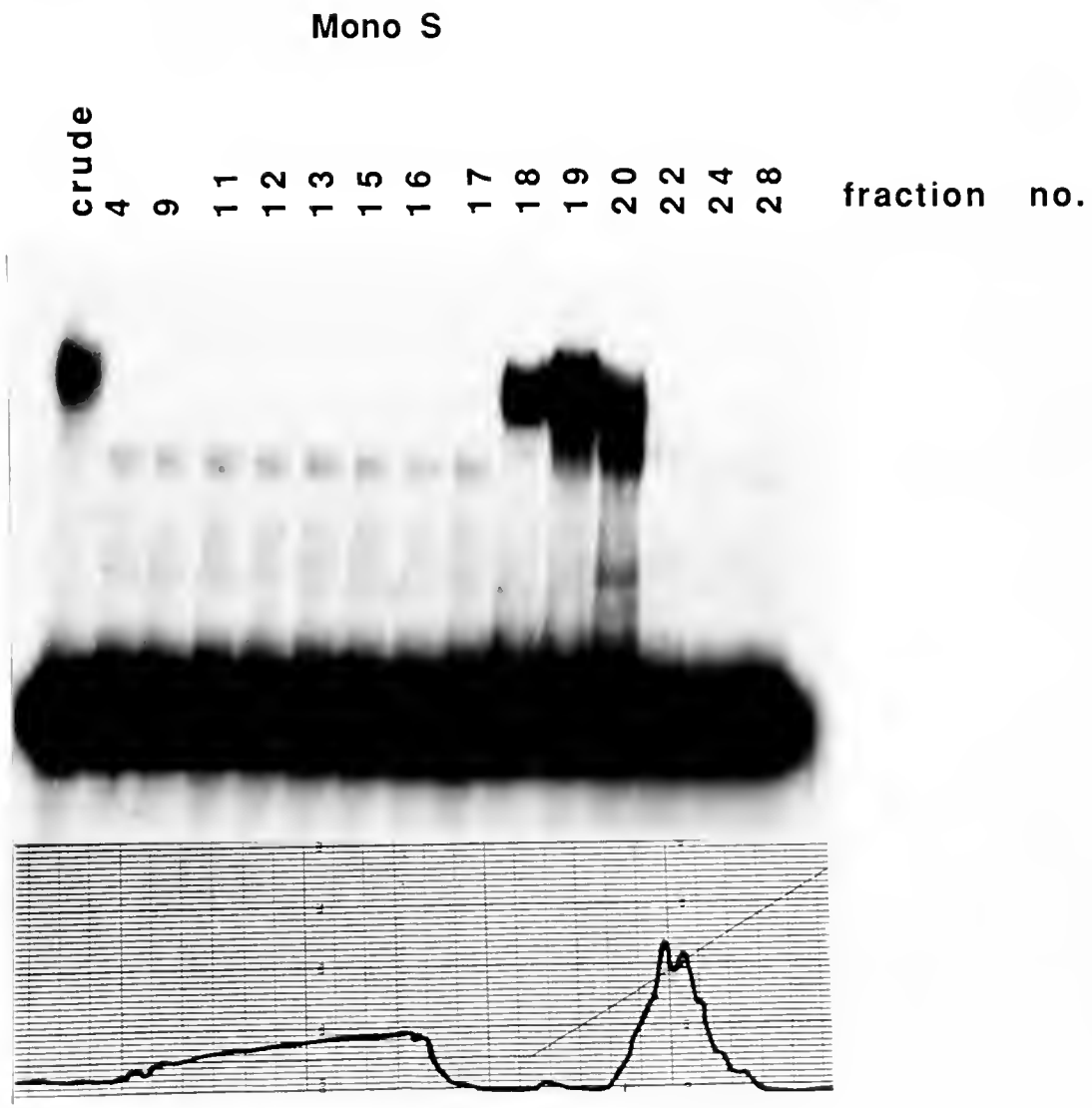


Figure 3-3. Mono-S elution profile. 780BP activity eluted in fractions (1 ml) 18,19, and 20. 1 µl of each fraction was added to the binding reaction and the gel retardation assay was performed as described in Materials and Methods. The chart below is a chromatogram of a typical Mono-S separation (A_{280}).

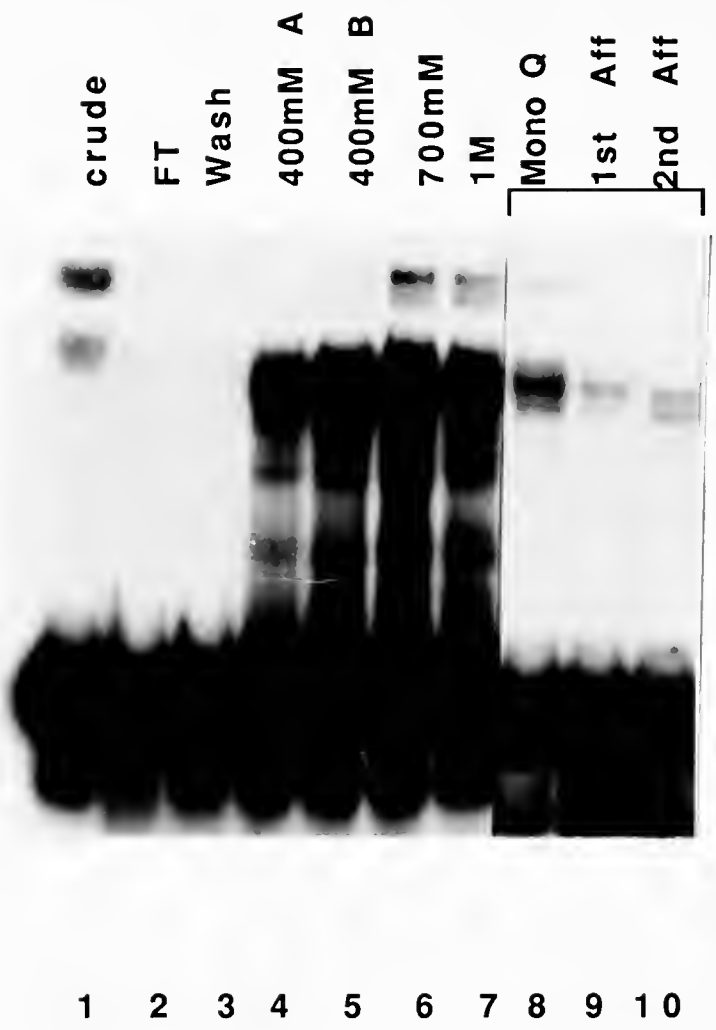


Figure 3-4. Affinity column elution profile. Lanes 2-7 contained binding activity that was retained on the affinity resin and eluted from 0.4 to 1 M KCl. Lanes 8-10 represent a different gel which demonstrates the one band-two band pattern seen throughout the purification process.

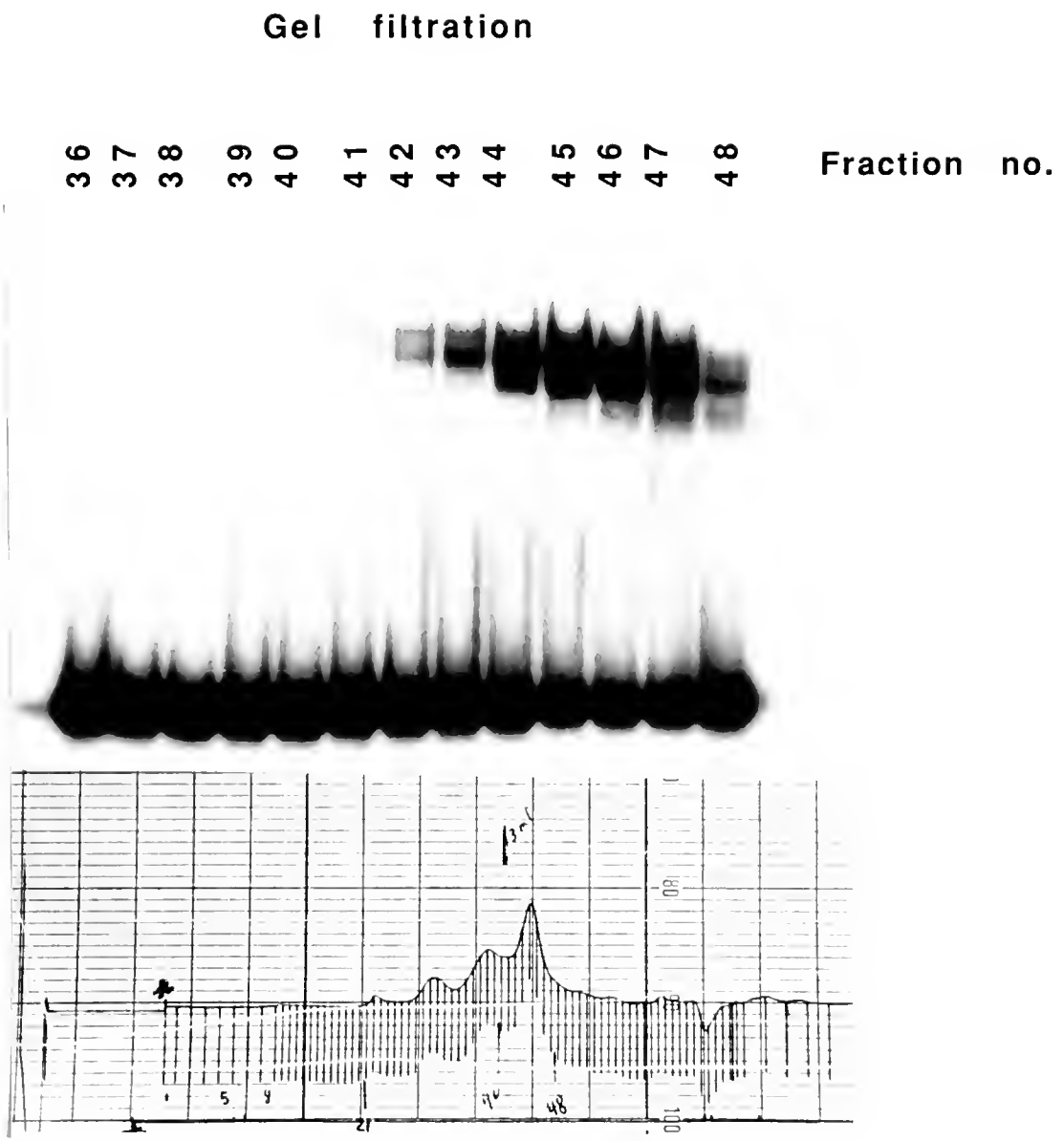


Figure 3-5. Superose 12 gel filtration profile. 50 ml of concentrated 1st affinity extract was loaded onto a Superose 12 column and eluted in 0.25 ml fractions. Fractions 1-55 were assayed for binding activity. Fractions with binding activity are indicated above. The chromatogram at the bottom is the UV profile of protein elution (A_{280}).

TABLE 3-1

	volume (microliters)	micrograms/ microliters	Total protein (micrograms)	°Bind Units/ microliter	Total Bind units	Specific Activity	Fold Purification
Crude Nuclear extract	15500	5.6	86800	1	15500	0.18	1
Heparin Sepharose	7000	4.2	29400	1.4	9800	0.19	1.8
Mono-Q	10000	0.75	7500	0.73	7300	0.73	5.4
Mono-S	3000	0.75	2300	0.67	2010	1	4.8
Affinity #1^	630	0.53	160	15	9450	59	328
Affinity #2	60		*1.6	4.7	352.5	220	*1222

^ = sum of divided samples
* = estimate

Table 3-1. Summary of purification of 780BP. °780 binding units = 1 fmol of DNA shifted.

resin for a second time. Eluted affinity pure protein was assayed, concentrated and stored for later use. A detailed account of yield, binding units determination, specific activity, and fold purification for each step of the purification scheme is given in Table 3-1. Fractionation over heparin-Sepharose resulted in the removal of 64% of the total protein. Mono-Q recovery was 9% of total protein. The purification of protein from Mono-Q to Mono-S was 33% which represented 3% of total starting protein. Specific activity and fold purification numbers for the affinity pure protein were derived from estimates of protein concentration from SDS-PAGE.

SDS-PAGE was used in combination with the silver staining technique to verify the relative molecular weight of the affinity pure protein, to monitor the purification process and to determine purity of the binding activity (Fig. 3-6). Many low and high mobility proteins were seen to disappear with each step in the purification scheme. Other proteins of variable molecular weight were also seen. Two bands in the 48-50 kD range were observed in the affinity fraction lane. Proteins in this molecular weight range were also observed when the affinity pure fraction was UV-crosslinked to the 780BP binding site (Fig. 3-7).

UV crosslinking experiments with crude nuclear extract have demonstrated the existence of two binding activities (48 kD, and 50 kD). When heparin pure protein and affinity pure protein were used in UV-crosslinking experiments along with

crude nuclear extract (Fig. 3-7B), two bands of 48 kD and 50kD were also observed. Non-specific activity seen with crude extract and with the heparin Sepharose fraction was not seen when affinity pure fractions were used. The UV crosslinking experiment was also used to answer the question of whether the proteins observed in the affinity pure lane of SDS-PAGE (Ag stained) are indeed the specific activity of interest. In this experiment, mutant oligonucleotides were used as competitors (Fig. 3-8). Since the mutants M142, M144, and M146 did not compete for specific binding activity it was predicted that the same specific banding pattern seen with crude nuclear extract and with affinity pure protein would be observed when these mutants were used as competitors. The results were quite surprising. Evidence has been presented previously that suggests that the doublet seen in gel retardation experiments and in UV crosslinking experiments with crude nuclear extract could be the result of protein degradation, and, thus, the same binding pattern should be observed in this experiment. The results from the UV crosslinking experiments with mutant competitors (Fig. 3-8) showed proteins of different molecular weights binding the wild type probe in the presence of each mutant oligonucleotide competitor. With no competitor (Fig. 3-8, lane 3), the typical banding pattern was seen with proteins of 50 and 52 kD. With mutant M142 as the competitor (Fig. 3-8, lane 5), two bands of lower mobility are seen (proteins of 57 and 56 kD). With mutant M144 as the competitor (Fig. 3-8,

lane 6), four bands of equal and lower mobility were seen (proteins of 57, 56, 52, and 50 kD). And, finally with mutant M146 as the competitor (Fig. 3-8, lane 7), two bands of the same mobility as in lane with no competitor were present (50 and 52 kD).

The 780BPE (5'-TTGAAAAATCAACGCT-3') shares a 6 out of 10 bp homology with the inverted repeat element (Fig.3-10) bound by the chicken ovalbumin upstream promoter transcription factor (COUP-TF). COUP-TF has been shown to be a member of the steriod receptor superfamily (Wang *et al.*, 1989). Analysis of COUP-TF binding specificity has shown that COUP-TF can bind to a series of naturally occurring elements with different spacing and orientation of the GGTCA motif repeat (Fig. 3-9A) (Cooney, 1992). Oligonucleotides containing the direct repeat and the inverted repeat with spacing variations from 0 to 12 bases were synthesized and used to test the extent of this promiscuous COUP-TF binding behavior. Because of the shared homology between the COUP-TF element and the 780BP, the hypothesis that the 780BP factor would bind the COUP-TF element was formulated. To test this hypothesis, an oligonucleotide containing the 2 bp direct repeat (Fig.3-10), to which COUP-TF had been shown to bind with high affinity, was synthesized and used as a competitor in gel retardation experiments with the 780BP and the 780BPE. Not only did the COUP oligomer compete for binding of 780BP activity, the efficiency of this competition indicated that it had greater affinity for 780BP than the wild type oligomer (Fig. 3-11).

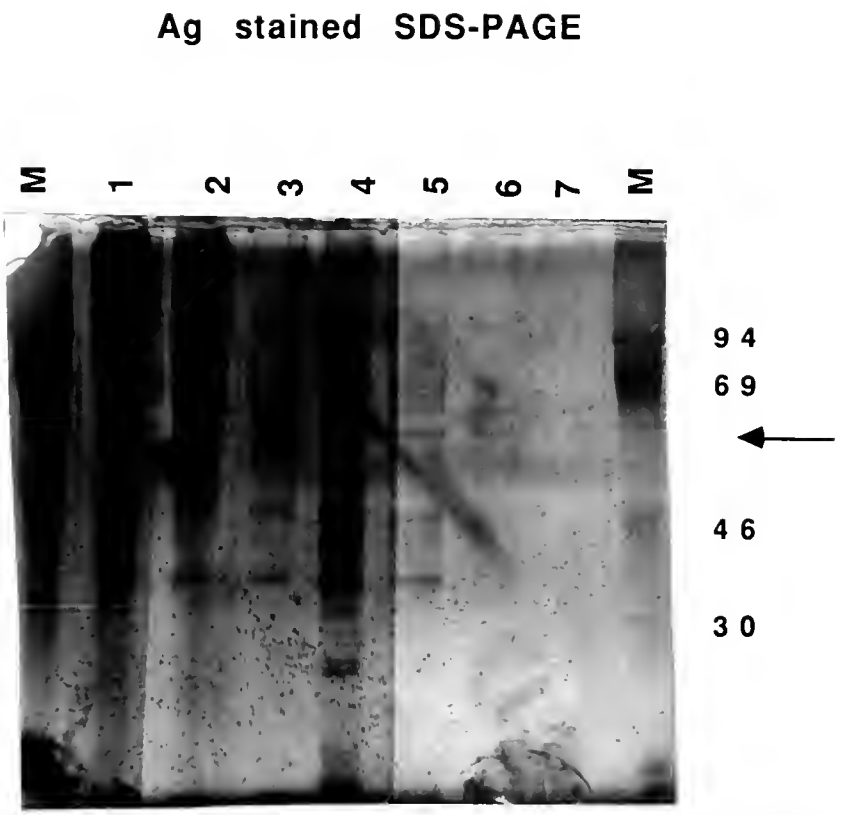


Figure 3-6. SDS-PAGE and Ag stained gel. Protein markers are in the lanes marked M. Lanes 1-4 contained 5 µg of crude extract, heparin fraction, Mono-Q fraction, and Mono-S fraction, respectively. Lane 5 contained 1.8 µg of column affinity fraction (different from batch affinity fraction), and lane 6 contained a batch affinity fraction that has been subjected to gel filtration. A single band of approximately 50 kD is visible in this lane (arrow). The band of lower mobility also seen in this lane is a gel artifact which can also be seen in lane7 which had no protein added.

Figure 3-7. UV-crosslinking with wild type probe, crude nuclear extract, heparin fraction, and affinity fraction. (A) Timed exposure to UV lamp. (B) Crude extract, C; heparin fraction, H; and affinity fraction, A. UV exposure time was 20 min. The second lane of each treatment in B contained a higher concentration of extract.

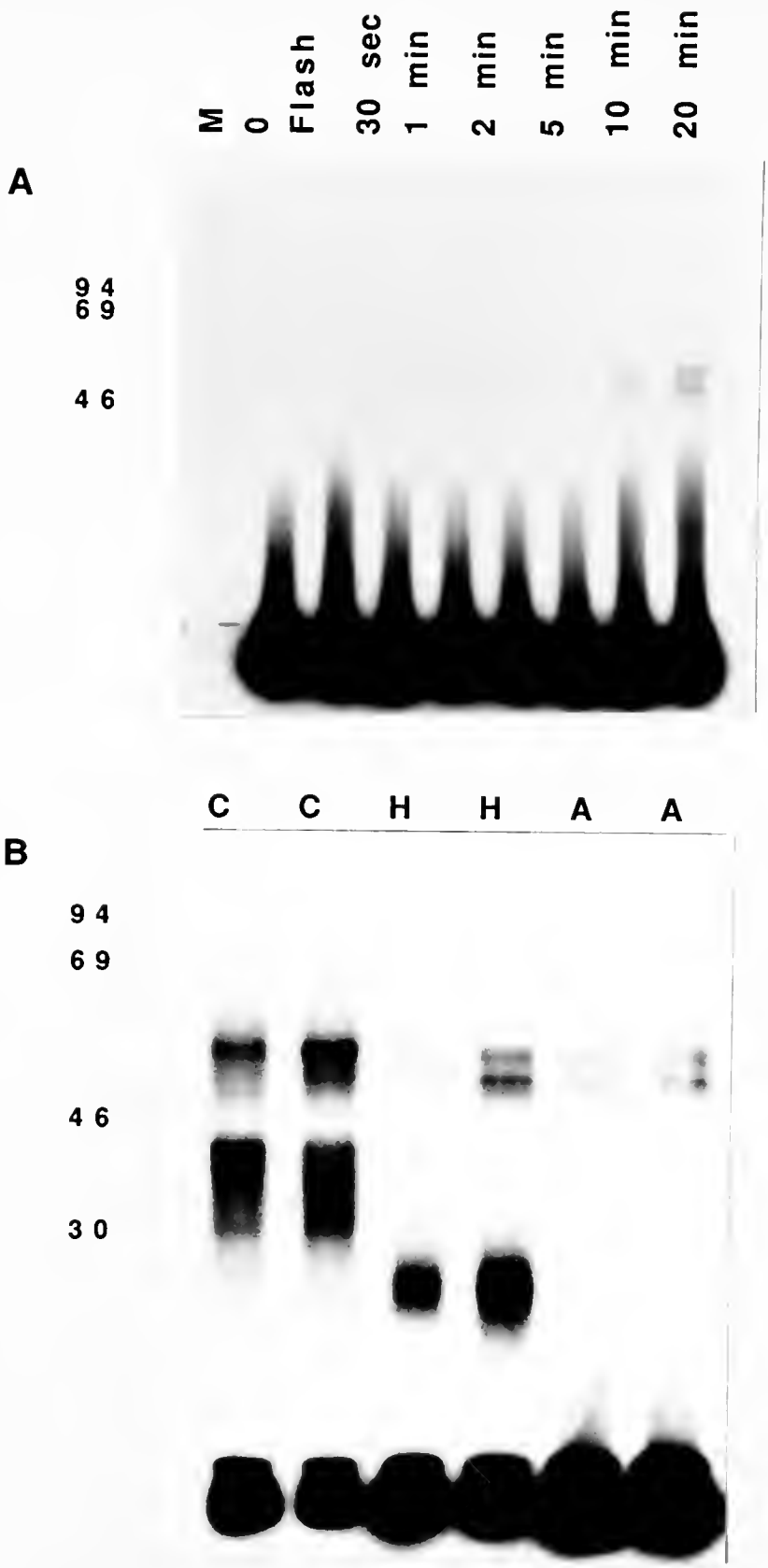
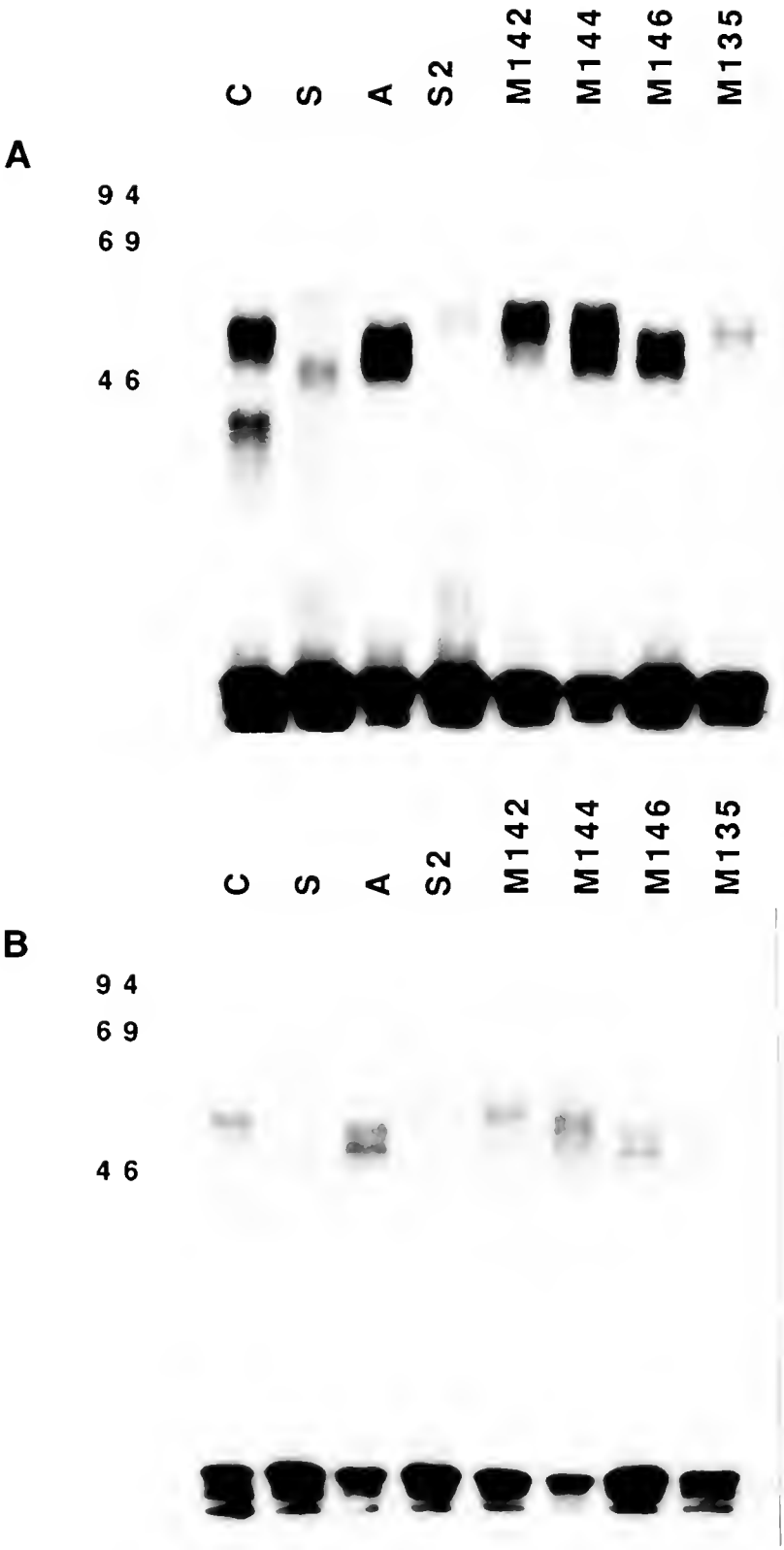

M
0
Flash
30 sec
1 min
2 min
5 min
10 min
20 min
A
94
69
46
C C H H A A
B
94
69
46
30

Figure 3-8. UV-crosslinking to wild type probe in the presence of mutant competitors. (A) represents a longer exposed version of (B) (17 hrs, 6 hrs respectively). Symbols: C indicates crude extract with no competitor; S., S2 indicates self competition; A indicates affinity fraction; and M indicates the mutant competitors. Affinity pure fraction used in all lanes except C and S 2. The mutants were competed at 100 fold molar excess.


A
C
S
A
S2
M142
M144
M146
M135
94
69
46
B
C
S
A
S2
M142
M144
M146
M135
94
69
46

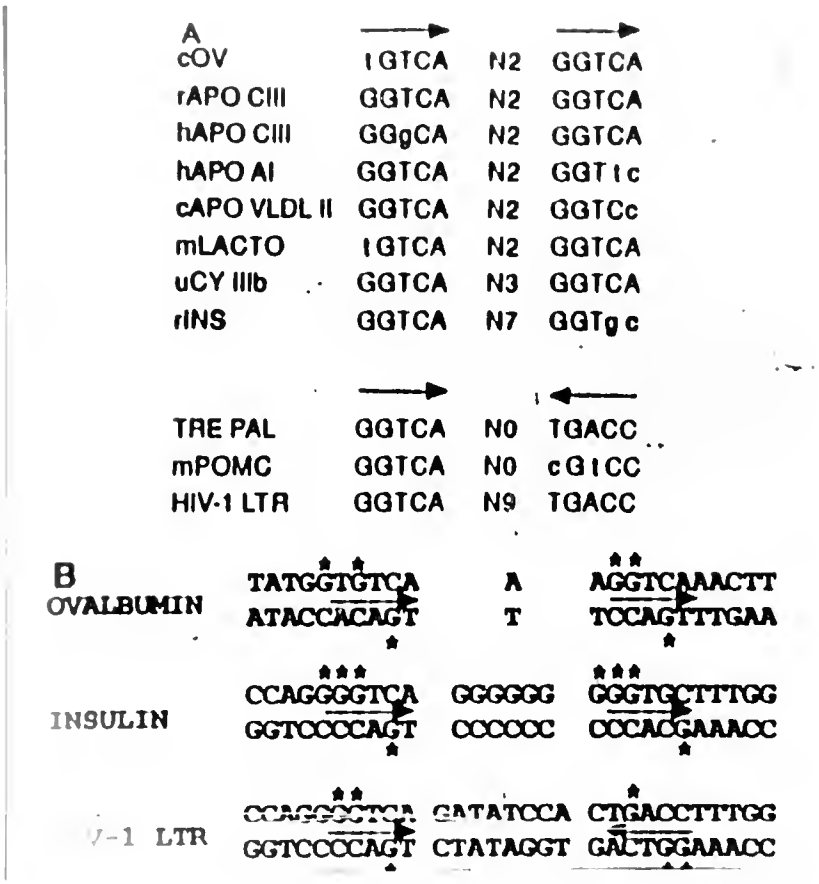


Figure 3-9. Sequences from Cooney *et al.* (1992) to which COUP-TF has been shown to bind. (A) The arrows indicate the orientation of the pentamer motif. Explanations of abbreviations are as follows; cOV, chicken ovalbumin; rAPO CIII, rat apolipoprotein CIII; hAPO cIII, human apolipoprotein CIII; hAPO AI, human apolipoprotein AI; cAPO VLDL II, chicken very low density apolipoprotein II; mLACTO. mouse lactoferrin; uCY IIIb, sea urchin actin IIIb; rINS, rat insulin II; TRE PAL, TRE palindrome; mPOMC, mouse pro-opiomelanocortin; HIV-1 LTR, human immunodeficiency virus type 1 long terminal repeat. N denotes the direct-repeat spacing, and N' denotes the palindrome spacing. (B) Summary of methylation interference data for binding of COUP-TF to the indicated probes.

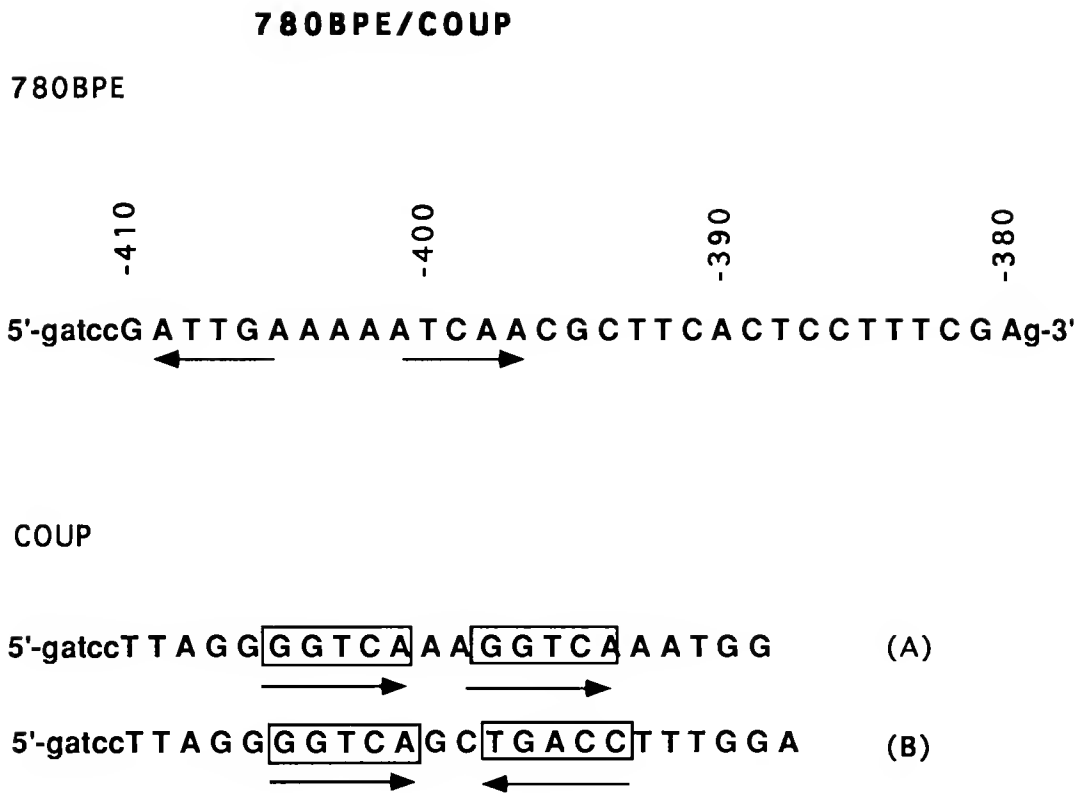


Figure 3-10. 780BPE/COUP-TF sequence comparison. COUP-TF probe A used as probe and in competition studies.

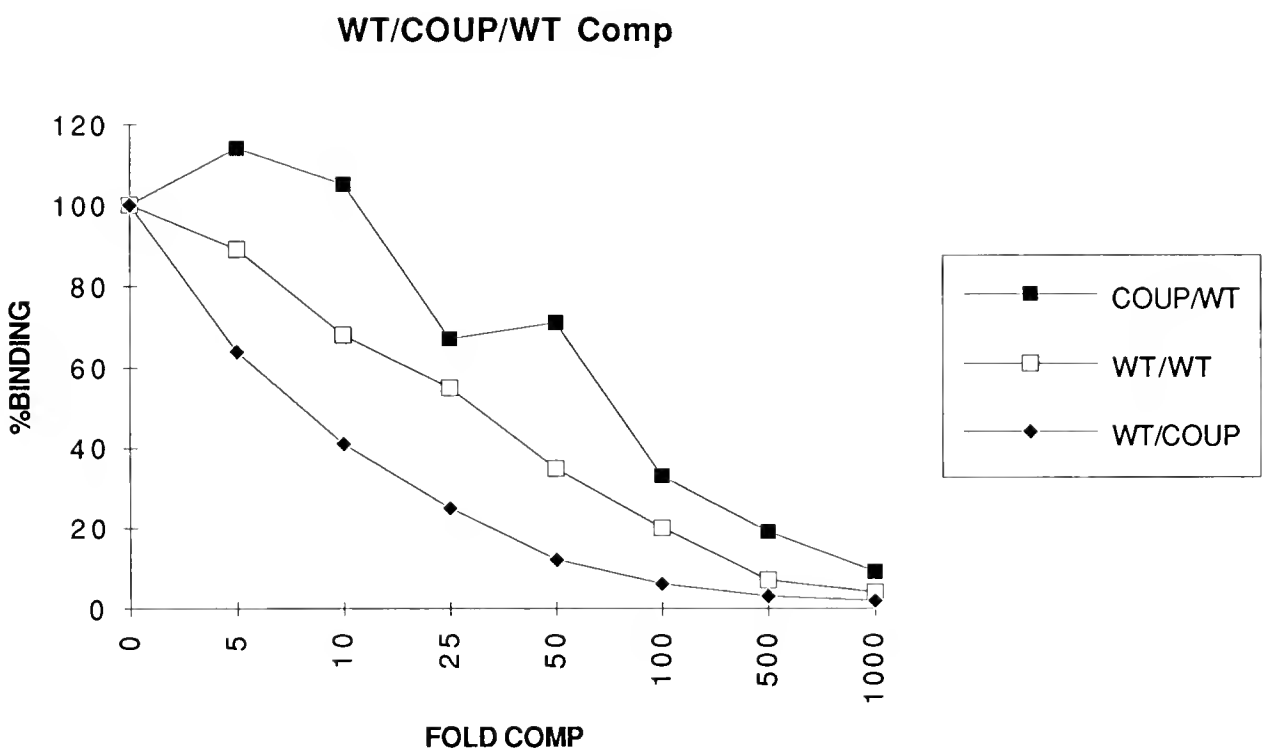


Figure 3-11. Summary of wild type competition with self and with the COUP-TF probe using cauliflower nuclear extract. COUP/WT indicates that COUP was the probe and WT was the competitor. The COUP-TF probe competed wild type better than wild type competed itself. Wild type did not compete COUP-TF probe very well.

When the COUP-TF element was used as a probe in gel retardation experiments, high affinity binding could be observed with a gel mobility similar to that obtained with the 780BPE probe. Compare lanes 1 and 2 in Fig. 3-12 with lanes 6 and 7. These results suggest that the binding activity that had been purified from cauliflower nuclear extract is a COUP-TF-like factor, or, perhaps, even a member of the steroid/thyroid hormone receptor superfamily.

COUP-TF has been purified from HeLa nuclear extract and cloned from HeLa cells (Wang *et al.*, 1987). Polyclonal antibodies have been raised for human COUP-TF and were shown to be specific by gel retardation "supershift" experiments (Wang *et al.*, 1989). In this experiments, the incubation of nuclear extract containing COUP-TF with the antibody prior to addition of the probe resulted in the formation of a lower mobility complex and the reduction or disappearance of the characteristic higher mobility COUP-TF-DNA complex (the supershift). Therefore, to determine whether this COUP-TF specific antibody would cause a supershift, when incubated with cauliflower inflorescence nuclear extract, a gel retardation experiment was designed and run. No supershift was observed which indicated that the COUP-TF antibody did not recognize any epitopes of the *780* binding activity (Fig. 3-12). These results were verified when a positive control was run on the same gel. A supershift was observed using HeLa cell nuclear extract from which COUP-TF was purified. Thus, 780BP appears to be different from COUP-TF; however,

the fact that it binds the hormone responsive element with high affinity must be significant.

Other common characteristics shared between COUP-TF and 780BP were explored. The dissociation rate of COUP TF-DNA complex was measured and a half-life of 2.4 min was reported (Tsai *et al.*, 1987). The dissociation rate of the 780BP-780BPE complex was analyzed and found to be less than 30 sec. However, when 780BP was complexed with the COUP-TF binding site, the half-life was approximately 2 min, well within the range of COUP-TF (Fig. 3-13).

All members of the nuclear receptor superfamily that have been cloned have been shown to have a DNA-binding domain with potential to form a zinc finger motif when complexed with DNA (Evans, 1988; Beato, 1989; Laudet *et al.*, 1992). EDTA and 1,10-phenanthroline are well known metal chelators and have been shown to inhibit protein-DNA complex formation between the *Xenopus* transcription factor TFIIIA and its cognate binding site (Hanas *et al.*, 1983), between histone H4 transcription factors H4TF-1 and H4TF-2 and their cognate site within the H4 promoter (Dailey *et al.*, 1987), and between the overexpressed DNA binding domain of the glucocorticord receptor and a cognate receptor binding site (Freedman *et al.*, 1988). Stability of the 780BP-780BPE complex was tested in the presence of increasing concentrations of these metal chelators (Fig. 3-14). EDTA concentration of 16 mM and 1, 10-phenanthroline concentration of 4 mM completely disrupted the formation of the 780BP-

780BPE complex. Several experiments have been performed for the purpose of reconstituting complex formation with the addition of Zn ions to a previously metal chelated reaction. Because the 780BP-780BPE complex has not been reconstituted to date, it is impossible to conclude that the inhibition of complex formation is the result of Zn chelation.

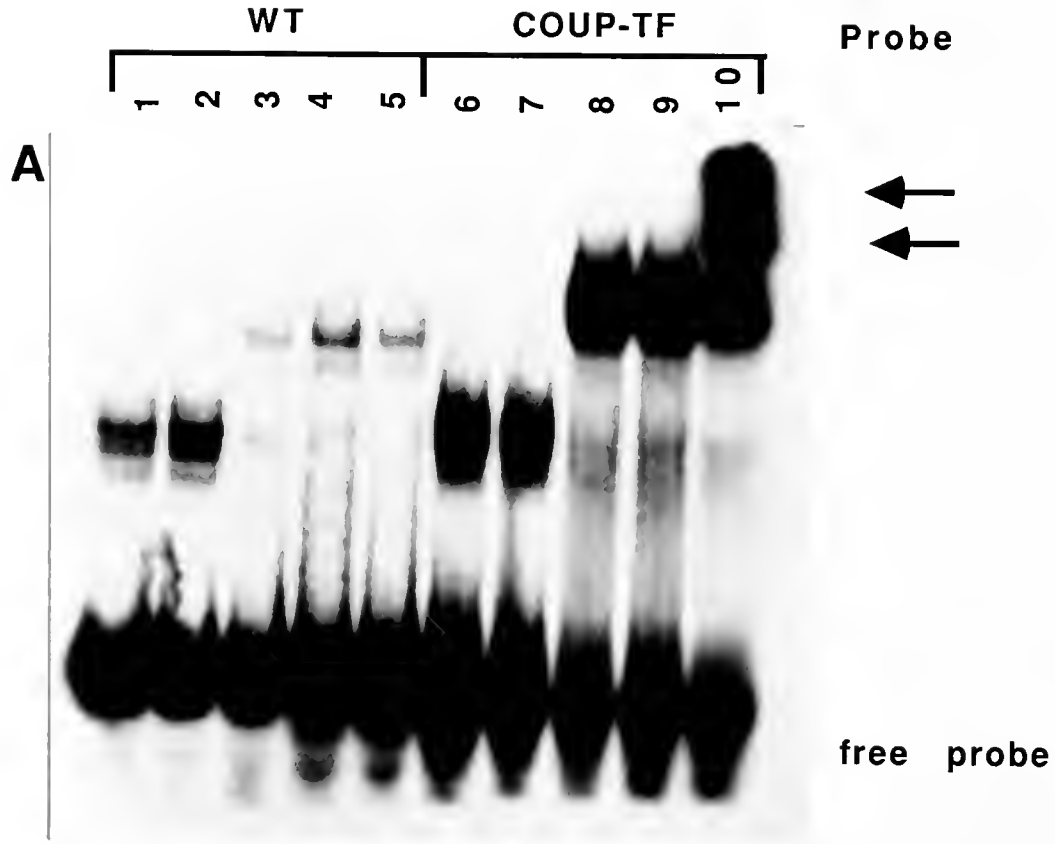


Figure 3-12. Supershift experiment. Probe: lanes 1-5, WT oligonucleotide; lanes 6-10, COUP-TF. Lanes 1, 2, 6, and 7 contained cauliflower nuclear extract. Lanes 3, 4, 5, 8, 9, and 10 contained Hela extract. COUP-TF antisera was added in lanes 5 and 10. The arrows indicate the supershift. Gel retardation assay conditions were the same as in Materials and Methods except that a 4% gel concentration was used.

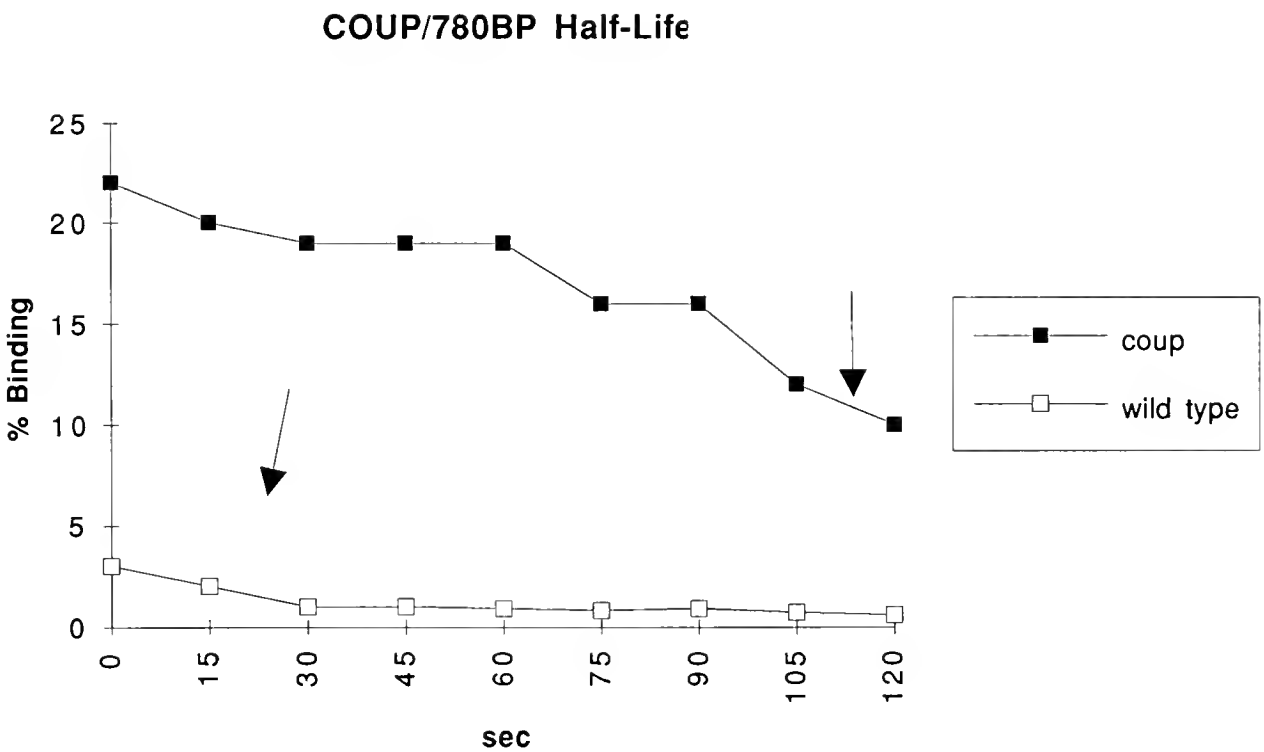


Figure 3-13. Summary of half-life experiments. Approximate half-life of each complex is indicated by an arrow.

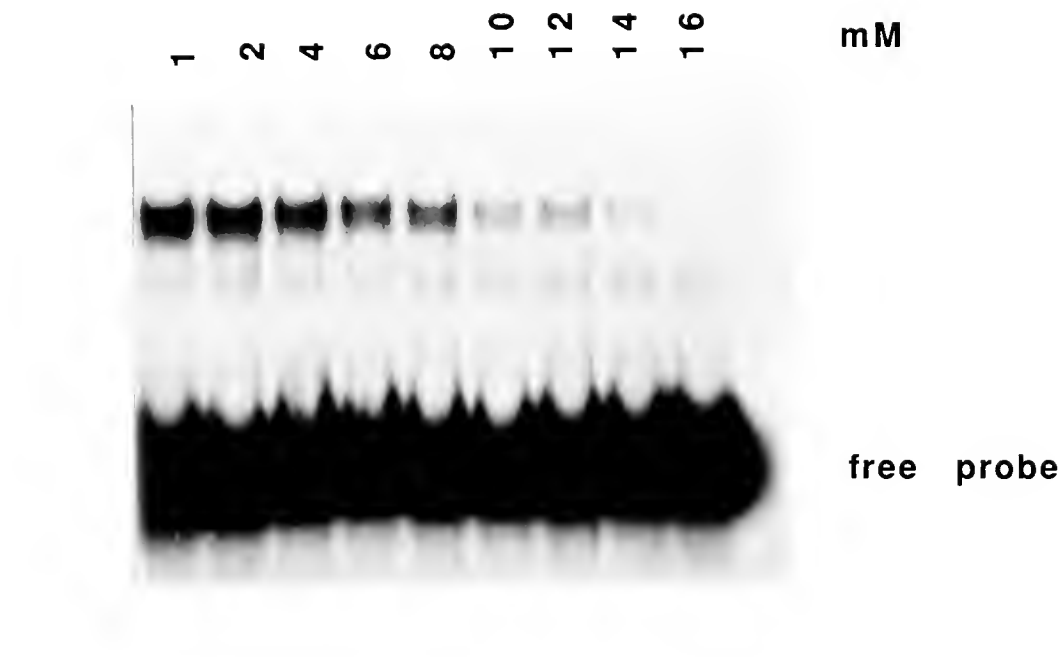


Figure 3-14. Effects of increasing concentrations of EDTA on formation of the 780BP-DNA complex. Gel retardation assays were run under conditions as described in the Materials and Methods of Chapter 2. The extract was incubated for 10 min with increasing concentrations of EDTA (Na_4) before the probe was added.

Discussion

A novel DNA binding protein activity (780BP) has been purified from nuclear extracts and partially characterized. The fold purification (1000 to 2000 fold) was within the range of that from reports of the purification of SBF-1 (Harrison, 1991) and from that reported for the purification of animal factors (Briggs *et al.*, 1986; Kadonaga & Tjian, 1986; Wu *et al.*, 1987). Two distinct protein-binding-activities (perhaps members of the same family or degradation products of the same protein) have been purified through a series of column chromatography steps including, heparin-Sepharose, FPLC Mono-Q and Mono-S ion exchange, FPLC Superose-12 gel filtration and sequence specific DNA affinity chromatography. Specific binding activity was monitored throughout the purification process with the gel retardation assay. The results of UV-crosslinking, SDS-PAGE (Ag stained), and Superose-12 gel filtration are consistent with the characterization of the 780BP activity purified as two monomeric proteins in the 48-50 kD range.

The low abundance of proteins in plant cells compared to animal cells, and the even lower abundance of transcription factors, make the purification of a plant transcription factor an arduous task. The difficulty of quantifying 780BP DNA binding activity in crude extracts renders information regarding protein recovery an estimate at best. Purification

for the purpose of obtaining quantities of protein sufficient for microsequencing and/or antibody production would require a five to ten fold increase of starting material.

During the process of determining the most efficient purification protocol for this binding activity, a variety of schemes were tested. The most efficient method is outlined in Fig. 3-14. Ammonium sulfate precipitation of crude nuclear extract, followed by passage over heparin-Sepharose, a Mono-Q, a Mono-S, and gel filtration columns, and DNA affinity chromatography (X2) is recommended. If the mutant oligonucleotide probes truly discriminate binding activity as is indicated in the UV-crosslinking experiment in Fig. 3-9B, then a mutant DNA affinity treatment may be useful.

For a plant DNA binding protein, 780BP was found to have a highly unusual characteristic in that it was shown to bind a steriod/thyroid nuclear receptor element (Cooney, 1992) with higher affinity than to 780BPE. When compared to an orphan member (COUP-TF) (Wang *et al.*, 1989) of this superfamily of nuclear receptor proteins, 780BP was found to share a number of similar characteristics of which binding to the nuclear receptor element being the most significant. Members of this superfamily are Zn-finger DNA binding proteins for which metal chelators such as EDTA and 1,10-phenanthroline have been shown to inhibit. When the 780BP was incubated with both of these metal chelators, the 780BP-780BPE complex was diminished or disappeared (Fig. 3-14). A final characteristic shared by COUP-TF and 780BP is the short

half-life of the protein-DNA complex. In a measure of stability of the protein-DNA complex involving COUP-TF and its cognate binding site, a half-life of 2.4 min was determined (Tsai *et al.*, 1987). The half-life of the complex formed between the 780BP and the nuclear receptor element was determined to be approximately 2 min, while the 780BP-780BPE complex demonstrated a half-life of less than 30 sec.

COUP-TF has been analyzed and characterized as a member of the steroid/thyroid hormone receptor superfamily. Perhaps 780BP will be the first plant member of this nuclear receptor superfamily to be identified.

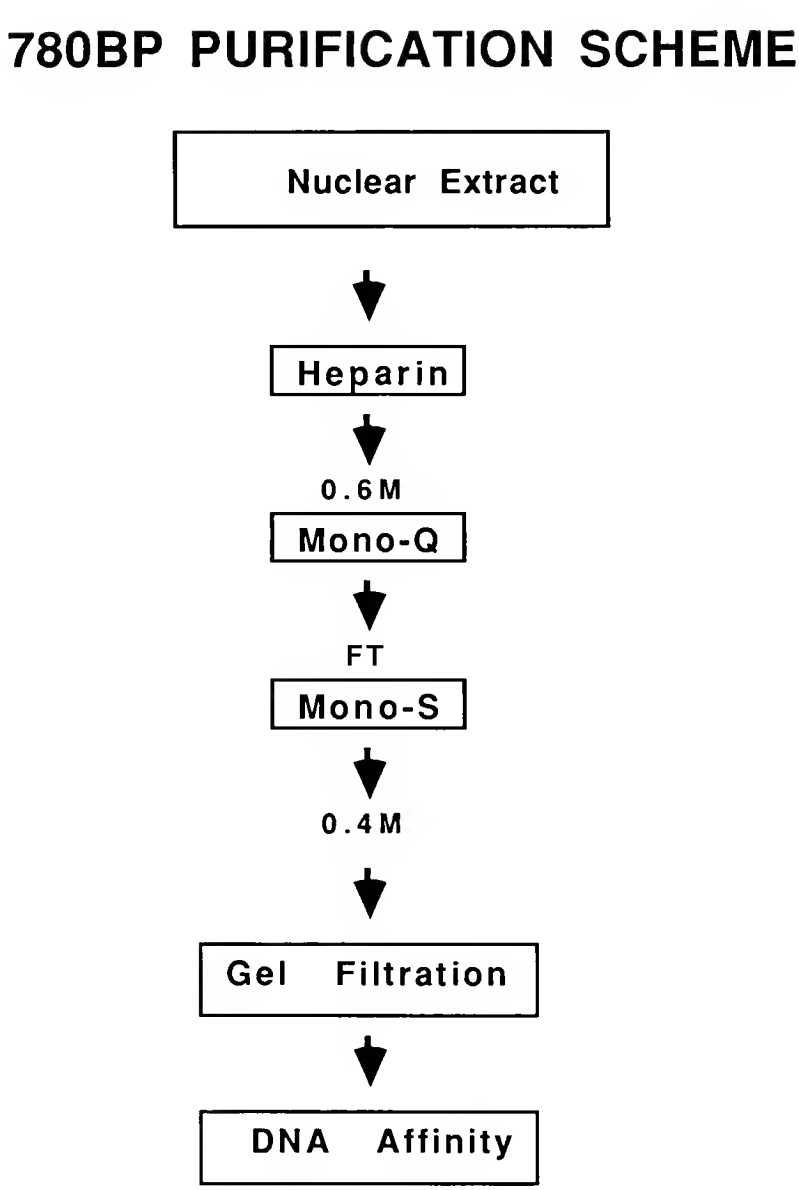


Figure 3-15. Suggested purification scheme for 780BP.

CHAPTER 4

SUMMARY AND DISCUSSION

The binding site of a nuclear protein (780BP) from cauliflower has been mapped to the position from -408 to -393 in the 5'-flanking sequences of the *780* gene of T-DNA. Double stranded oligonucleotides synthesized with sequences homologous to a region in the *780* upstream activator region (Bruce & Gurley, 1987) were used as probes in gel retardation assays with nuclear extracts to demonstrate DNA-protein interactions. Kinetic competition studies, methylation interference assays, and $KMnO_4$ interferences assays were used to defined specific bases required for binding. Protein binding activity specific for this binding site (780BPE) was purified and characterized with a series of steps which included ion exchange, gel filtration, heparin affinity chromatography, and sequence specific DNA affinity chromatography.

The role of the 780BPE as a transcription regulatory element has not been shown directly, but is suggested by the demonstration of its specific interaction with a nuclear protein. Further evidence in support of this protein-binding site being a transcription regulatory element can be seen in the deletion studies by Bruce and Gurley (1987) and in the

site-directed mutagenesis studies by O'Grady (1993). Bruce and Gurley showed that when a 31 bp region containing this element was deleted, transcription from the *780* promoter was decreased by 50%. In addition, O'Grady demonstrated that substitution mutations at two unique sites within this region significantly influenced transcription in sunflower tumors. One mutation resulted in an increase in transcription and the other resulted in a decrease. Though strongly implicated as containing at least one transcriptional regulatory element, convincing evidence to support the designation of the 780BPE as a promoter elememt can only come from experiments in which the 780BPE incorporated into a chimeric construction with a truncated promoter and a reporter gene reduces or enhances transcriptional activity.

The DNA-binding protein 780BP was shown with the use of SDS-PAGE, UV-crosslinking, and gel filtration to be a monomer in solution and to have a molecular weight in the range of 48 to 50 kD. Since the gel filtration experiment and the UV-crosslinking experiment indicated similar molecular weights, it was originally concluded that this protein binds DNA as a monomer. However, this conclusion should be modified since it was not shown that the UV crosslinking was sufficient to covalently attach two proteins to the same DNA probe. Therefore, the 780BP has been shown to exist in solution as a monomer, but it is not clear how many 780BPs are required to bind the 780BPE. The two bands (protein-DNA complexes)

observed in some gel retardation assays and after UV-crosslinking probably resulted from protein degradation.

A simple model (Model 1) to explain the formation of protein-DNA complexes of two different mobilities is shown in Fig. 4-1. This conclusion is most evident after considering all of the data. In the competition studies seen in Figs. 2-5, 2-6, 2-7, and 2-11, the extracts used were from once thawed preparations and a single band is clearly evident. During the purification process, an especially convincing result can be seen in Fig. 3-4, lanes 7-10, where one band predominates in the Mono-Q and first affinity fractions while two distinct bands are observed in the second affinity fraction. In preparation of the second affinity fraction, the protein was subjected to far more handling and thus was more likely to be subject to protein degradation.

Although protein degradation seems to be the most plausible explanation for the presence of two complexes of different mobilities, other explanations are possible. For example, one interpretation of these results is that protein-protein interactions are occurring between the 780BP and another non-DNA-binding protein (Model 2 in Fig. 4-2). This concept was dispelled by the results of the urea, NP-40, and Ca^{2+} disruption experiments in Fig. 2-14. The doublet persisted in the presence of high concentrations of urea and other reagents known to disrupt protein-protein interactions.

One experiment, the UV-crosslinking experiment with mutant oligonucleotides as competitors shown in Fig. 3-8,

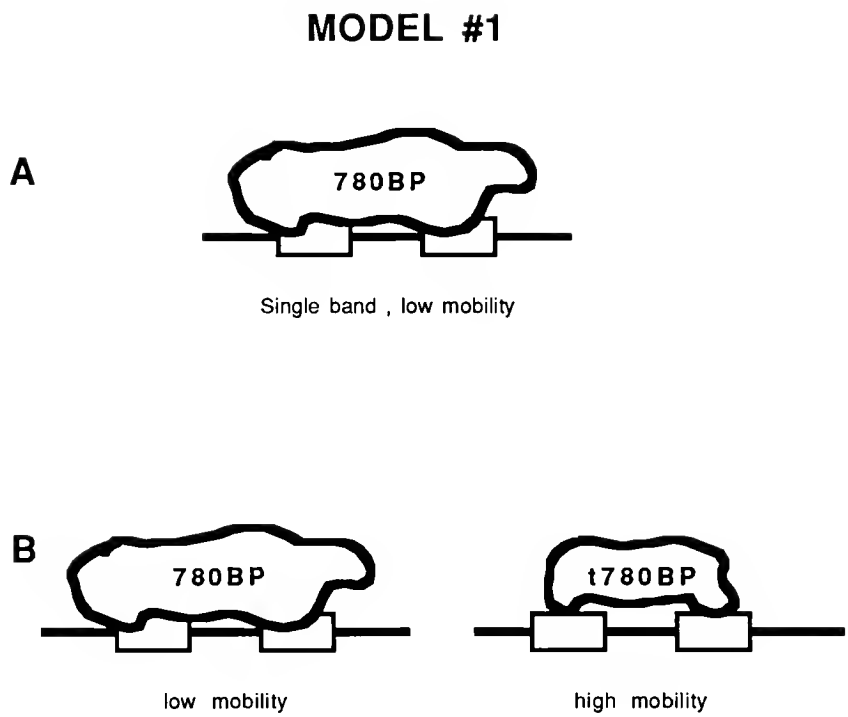


Figure 4-1. Model #1. A simple model to explain the banding pattern observed in the gel retardation assays. (A) A single protein binds 780BPE. (B) Under conditions of prolonged handling and after freezing and thawing, protein degradation results in the formation of two bands. Truncated 780BP (t780BP) represents a degraded form of the 780BP that is still able to bind DNA.

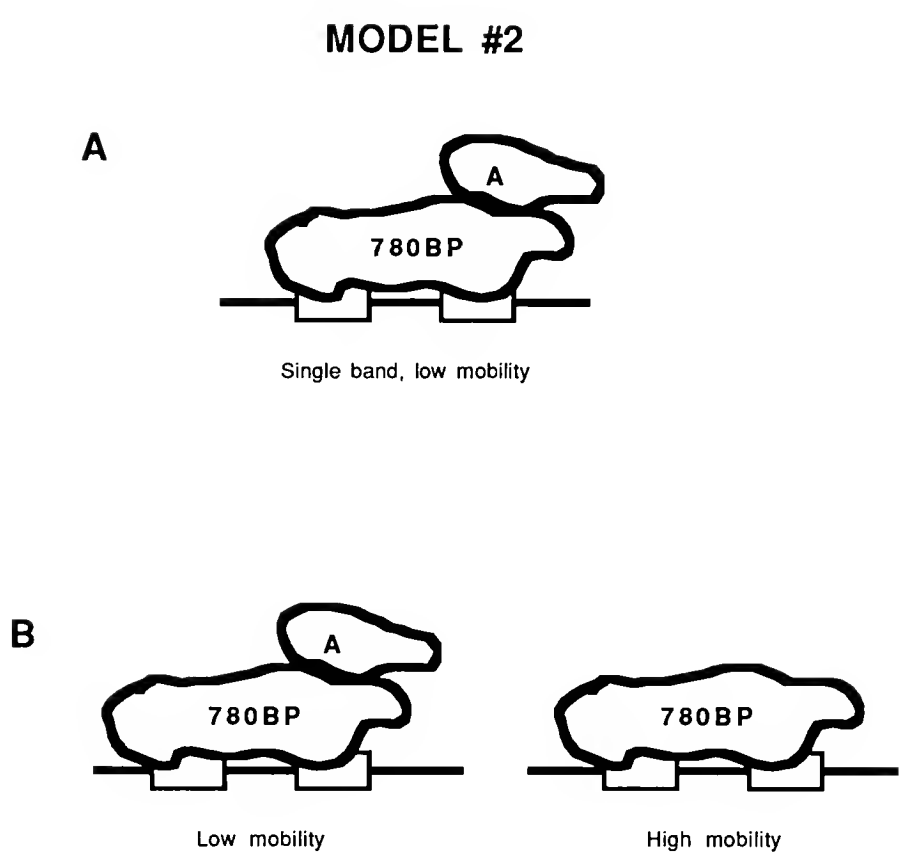


Figure 4-2. Model #2. (A) 780BP is bound to a smaller protein A through protein-protein interactions. These two closely associated proteins co-purify and bind the 780BPE through specific interactions with the 780BP. (B) Under certain conditions, partial dissociation occurs. The doublet banding pattern results when a mixture of the 780BP alone and complexed with protein A bind the 780BPE.

suggests a third explanation for this doublet banding pattern. The results show the binding activity changing to two bands of higher mobility and to four bands (two of higher mobility and two of the same mobility) when crosslinking occurred in the presence of unlabeled mutant oligomers M142 and M144, respectively. With mutant M146 the overall intensity of the crosslinked protein bands were decreased, but the banding pattern was the same as when no competitor was added. The complexity of binding activities could have resulted from the interactions of a family of proteins that bind with different affinities depending on the purity of the protein preparation. This possibility seems unlikely, especially since the doublet can be seen in crude extracts that have been subjected to numerous freezing and thawing cycles. Perhaps this family of proteins is present in crude extract, but the UV-crosslinking experiment with mutant competitors was not performed and thus the family not detected.

There is yet another explanation that could explain the banding phenomenon observed in the experiment in Fig. 3-8. Different conformations of one protein(s) could be binding overlapping sequences with different affinities. A similar observation was made when COUP-TF was shown to bind HREs of variable spacing between direct and inverted repeats (Cooney *et al.*, 1992). Proposed models for both the family of proteins and the conformational change hypotheses can be seen in Fig. 4-3. Each of the proteins described in Fig. 4-3

exhibits higher affinity for specific bases within the two half-sites of 780BPE. For purposes of this discussion, the two half-sites are further divided into sub-sites 1A, 1B, 2A and 2B. Protein 1 (p1) prefers sub-sites 1B and 2A. Protein 2 (p2) prefers subsites 1B and 2B. Protein 3 (p3) prefers sites 1A and 2B. Consequently, M142 competes away p1 and p2 leaving p3 which produces lower mobility complexes. M144 competes away p1 and leaves p2 and p3, and M146 competes away p2 and p3 and leaves p1 which produces bands of the same mobility as p2.

Perhaps the most significant result of this work is the identification of a previously uncharacterized plant protein that interacts with DNA sequences from a T-DNA promoter. The OCSTF (Tokuhisa *et al.*, 1990) and OCSBF (Singh *et al.*, 1990), which may prove to be the same factor, are the only other plant transcription factors identified that interact specifically with an element of a T-DNA promoter. The existence of transcription factors in plants that interact with promoter elements of genes introduced by plant pathogens such as *Agrobacterium* implies that there are cellular counterparts of these elements in the promoters of plant genes. Thus, a logical extension of this work would be to search the gene banks for sequence homology to the 780BPE.

780BP was shown to have some similarity to COUP-TF (Wang *et al.*, 1989), an orphan receptor member of the steroid/thyroid receptor superfamily (Evans, 1988). Members of this nuclear receptor family have been studied and

reviewed extensively (Beato, 1989; Evans, 1988; Green & Chambon, 1988). This superfamily of nuclear receptors includes a total of 32 proteins among which are the steroid, thyroid hormone, retinoic acid, and vitamin D_3 receptor proteins (Laudet *et al.*, 1992). Based upon structural and functional similarities between these cloned proteins, the amino acid sequences of the members of this superfamily have been divided into four domains: domain 1 comprised of the N-terminal region which is hypervariable and shows low homology between receptor protein; domain 2 or the DNA-binding domain which contains two Zn-finger binding motifs; domain 3 which is involved in the transcriptional activation, nuclear translocation and heat shock protein binding; and domain 4 which is the ligand binding domain (Beato, 1989; Evans, 1988; Power *et al.*, 1992). COUP-TF is classified as an orphan member of this nuclear receptor superfamily because, even though it has all of the major structural characteristics of the classical receptors, no ligand has been identified.

The most significant characteristic shared by the 780BP and the COUP-TF is the ability to bind a mammalian hormone receptor element with high affinity. This finding strongly suggests that these two proteins have similar DNA binding domains. Homology in the DNA-binding domain of nuclear receptor proteins has been used as a major characteristic for studying the evolutionary relationship between members of the steroid/thyroid receptor superfamily (Laudet *et al.*, 1992). Laudet *et al.* (1992) constructed and compared the

phylogenetic trees derived from the DNA binding domain and from the transcriptional regulation and dimerization domain. Both phylogenetic trees demonstrated clearly that all nuclear receptors share common ancestry. The significance of the very strong conservation of DNA binding domains in mammals is further supported since the screening of cDNA libraries for nuclear receptors with DNA probes derived from the DNA binding domain of previously cloned nuclear receptor proteins has resulted in a total of 32 nuclear receptor proteins cloned to date (Laudet *et al.*, 1992; and references therein).

The hypothesis that 780BP will prove to be another member of the steroid/thyroid hormone receptor is not as novel as it may seem. Steroids and retinoids share a common biosynthetic pathway with the plant hormones gibberellic acid and abscissic acid. These hormones are derived by assembly of isoprene units and, because of this, are grouped as terpenoid hormones (Moore, 1990). Moore (1990) suggests that the terpenoids may be ligands of yet unidentified hormone receptors. Other relevant findings are reports of the isolation and physiological effects of steroid hormones in plants (reviewed in Heftmann, 1977; McMorris, 1978; Geuns, 1978; Hewitt *et al.*, 1980; Slama, 1980). Mitchell *et al.* (1970) extracted an active growth substance from the pollen of rape (*Brassica napus*) and alder (*Alnus glutinosa*), the structure of which was elucidated later and found to be that of a steroid hormone brassinolide (Grove *et al.*, 1979). This compound resembles ecdysteroids found in *Drosophila* and was

tested extensively in bioassays for the plant growth substances (auxins, gibberellins, and cytokinins) and found not to behave exclusively as any one group (Yopp *et al.*, 1979; Yopp *et al.*, 1981; Mandava *et al.*, 1981). In addition, other studies have documented the effects of corticosteroids on the elongation growth and lateral root formation on certain plants, and the involvement of a receptor protein has been postulated (Geuns, 1983).

The role of the 780BPE and that of the 780BP is open to much speculation. The 780BP could prove to be a nuclear receptor for a biologically active compounds such as auxin, cytokinin, gibberellic acid, abscissic acid, or various metabolites of each. Further extension of this line of reasoning suggests that it may activate or repress plant gene expression in a ligand dependent or independent manner, or, perhaps, be induced by a second messenger.

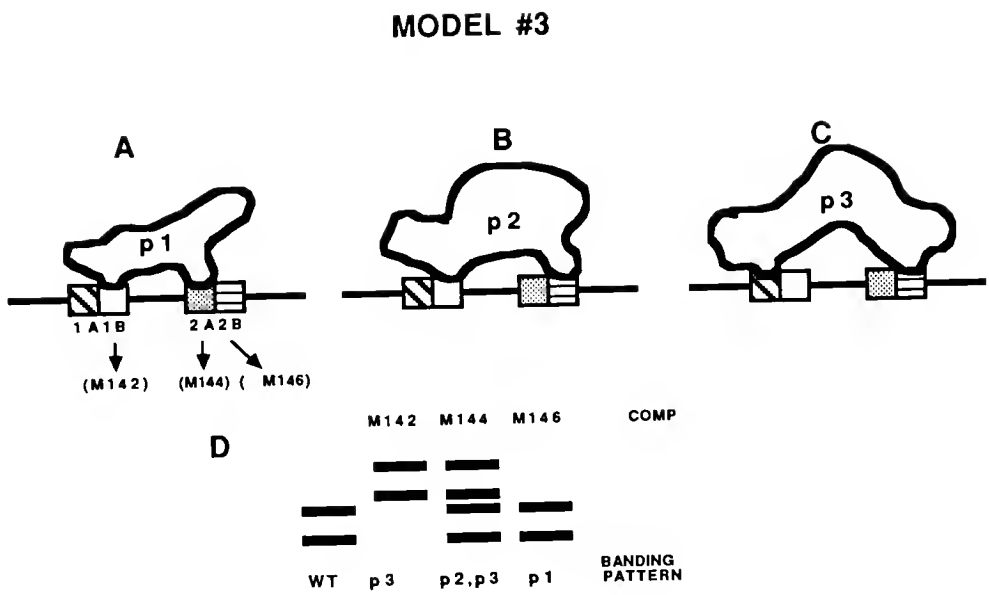


Figure 4-3. Model #3. This model is applicable for two conditions: (1) a family of proteins is able interact at the same or overlapping sites resulting in two protein-DNA complexes of differing mobilities; or (2) a single protein is capable of conformational change. Complex A demonstrates the contact of protein 1 (p1) with bases in each half site. Complex B demonstrates a different contact pattern for protein 2 (p2). Complex C demonstrates yet another contact pattern for protein 3 (p3). The location of each mutation is indicated (M142, M144, and M146). (D) Replica of banding pattern observed in UV-crosslinking experiment with mutant competitors. M142 competes away p1 and p2 leaving p3 which produces lower mobility bands. M144 competes away p1 and leaves p2 and p3. M146 competes away p2 and p3 and leaves p1 which produces bands of the same mobility as p2.

REFERENCES

An, G., Ebert, P. R., Yi, B. Y. & Choi, C.-H. (1986) *Mol. Gen. Genet.* 203, 245-250.

An, G., Costa, M. A., Mitra, A., Ha, S.-B. & Marton, L. (1988) *Plant Physiol.* 88, 547-552.

An, G., Mitra, A., Choi, H. K., Costa, M., An, K., Thornburg, R. W. & Ryan, C. A. (1989) *Plant Cell* 1, 115-122.

An, G., Costa, M. A. & Ha, S.-B. (1990) *Plant Cell* 2, 225-233.

Arias, J. A., Dixon, R. A. & Lamb, C. J. (1993) *Plant Cell* 5, 485-496.

Armstrong, G. A., Weisshaar, B. & Hahlbrock, K. (1992) *Plant Cell* 4, 525-537.

Bandyopadhyay, R. S., Bruce, W. B. & Gurley, W. B. (1989) *J. of Biol Chem.* 264, 19399-19406.

Banerji, J., Rusconi, S. & Schaffner, W. (1981) *Cell* 27, 299-308.

Barker, R. F., Idler, K. B., Thompson, D. V. & Kemp, J. D. (1983) *Plant Mol. Biol.* 2, 335-350.

Barros, M. D., Czarnecka, E. & Gurley, W. B. (1992) *Plant Mol. Biol.* 19, 665-675.

Baumann, G., Raschke, E., Bevan, M. &, Schoffl, F. (1987) *EMBO J.* 6, 1161-1166.

Beato, M. (1989) *Cell* 56, 335-344.

Benfey, P., Ren, L. & Chua, N.-H. (1989) *EMBO J.* 8, 2195-2202.

Benfey, P. N. & Chua, N.-H. (1990) *Science* 250, 959-966.

Benfey, P. N., Ren, L. & Chua, N.-H. (1990) *EMBO J.* 9, 1685-1696.

Benoist & Chambon (1981) *Nature* 290, 304-310.

Bevan, M. W. & Chilton, M.-D. (1982) *Annu. Rev. Genet.* 16, 357-384.

Bouchez, D., Tokuhisa, J. G., Llewellyn, D. J., Dennis, E. S. & Ellis, J. G. (1989) *EMBO J.* 8, 4197-4204.

Breathnach, R. & Chambon, P. (1981) *Ann. Rev. Biochem.* 50, 349-383.

Brenowitz, M., Senear, D. F., Shea, M. A. & Ackers, G. K. (1986) *Methods in Enzymology* 130 , 132-181. Academic Press.

Briggs, M. R., Kadonaga, J. T., Bell, S. P. & Tjian, R. (1986) *Science* 234, 47-52.

Brown, J. W. S., Wandelt, C. & Feix, G. (1986) *Eur. J. Cell. Biol.* 42, 161-170.

Bruce, W. B. & Gurley, W. B. (1987) *Mol. Cell. Biol.* 7, 59-67.

Bruce, W. B., Bandyopadhyay, R. & Gurley, W. B. (1988) *Proc. Natl. Acad. Sci. USA.* 85, 4310-4314.

Bucher, P. (1990) *J. Mol. Biol.* 212, 563-578.

Bustos, M. M., Guiltinan, M. J., Jordano, J., Begum, D., Kalkan, F. A., & Hall, T. C. (1989) *Plant Cell* 1, 839-853.

Chilton, M.-D., Drummond, M. H., Merlo, D. J., Sciaky, D., Montoya, A. L., Gordon, M. P. & Nester, E. W. (1977) *Cell* 11, 263-271,

Chilton, M.-D., Saiki, R. K., Yadav, N., Gordon, M. P. & Quetier, F. (1980) *Proc. Natl. Acad. Sci. USA*. 77, 4060-4064.

Chilton, M. D., Tepfer, D., Petit, A., David, C., Casse-Delbart, F. & Tempe, J. (1982) *Nature* 295, 432-434.

Chodosh, L. A., Carthew, R. W. & Sharp, P. A. (1986) *Mol. Cell. Biol.* 6, 4723-4733.

Cohen, R. B., Sheffery, M. & Kim, C. G. (1986) *Mol. Cell Biol.* 6, 821-832.

Comai, L., Moran, P. & Maslyar, D. (1990) *Plant Mol. Biol.* 15, 373-381.

Cooney, A. J., Tsai, S. Y., O'Malley, B. W., & Tsai, M.-J. (1992) *Mol. Cell. Biol.* 12, 4153-4163.

Cramer, C. L., Edwards, K., Dron, M., Liang, X., Dildine, S. L., Bolwell, G. P., Dixon, R. A., Lamb, C. J. & Schuch, W. (1989) *Plant Mol. Biol.* 12, 367-383.

Czarnecka, E., Key, J. L., & Gurley, W. B. (1989) *Mol. Cell. Biol.* 9, 3457-3463.

Czarnecka, E., Fox, P. & Gurley, W. B. (1990) *Plant Physiol.* 94, 935-943.

Czarnecka, E., Ingersoll, J. C. & Gurley, W. B. (1992) *Plant Mol. Biol.* 19, 985-1000.

Dailey, L., Roberts, S. B. & Heintz, N. (1987) *Mol. Cell. Biol.* 7, 4582-4584.

Datta, N. & Cashmore, A. R. (1989) *Plant Cell* 1, 1069-1077.

De Greve, H., Dhaese, P., Seurink, J., Lemmers, M., Van Montagu, M. & Schell, J. (1983) *J. Mol. Appl. Genet.* 1, 499-511.

de Pater, B. S., Klinkhamer, M. P., Amesz, P. A., de Kam, R. J., Memelink, J., Hoge, J. H. C. & Schilperoot, R. A. (1987) *Nucl. Acids Res.* 15, 8267-8281.

de Pater, B. S., de Kam, R. J., Hoge, J. H. C. & Schilperoort, R. A. (1987) *Nucl. Acids Res.* 15, 8267-8281.

Dehesh, K., Bruce, W. B. & Quail, P. H. (1990) *Science* 250, 1397-1399.

DeLisle, A. & Ferl, R. J. (1990) *Plant Cell* 2, 547-557.

DiRita, V. J. & Gelvin, S. B. (1987) *Mol. Gen. Genet.* 207, 233-241.

Dixon, R. A., Dey, P. M. & Lamb, C. J. (1983) *Adv. Enzymol. Relat. Areas Mol. Biol.* 55, 1-135.

Dron, M., Clouse, S. D., Dixon, R. A., Lawton, M. A. & Lamb, C. J. (1988) *Proc. Natl. Acad. Si. USA.* 85, 6738-6742.

Dymock, D., Risiott, R., de Pater, S., Lancaster, J., Tillison, P. & Ooms, G. (1991) *Plant Mol. Biol.* 17, 711-725.

Dynan, W. S. & Tjian, R. (1985) *Nature* 316, 774-778.

Ebert, P. R., Ha, S.-B. & An, G. (1987) *Proc. Natl. Acad. Sci. USA.* 84, 5745-5749.

Ellis, J. G., Ryder, M. H. & Tate, M. E. (1984) *Mol. Gen. Genet.* 195, 466-473.

Ellis, J. G., Llewellyn, D. J., Walker, J. C., Dennis, E. S. & Peacock, W. J. (1987) *EMBO J.* 6, 3203-3208.

Ellis, J. G., Llewellyn, D. J., Dennis, E. S. & Peacock, W. J. (1987) *EMBO J.* 6, 11-16.

Evans, R., M (1988) *Science* 240, 889-895.

Everett, R. D., Baty, D. & Chambon, P. (1983) *Nucl. Acids. Res.* 11, 2447-2464.

Fang, R.-X., Nagy, F., Sivasubramaniam, S. & Chua, N.-H. (1989) *Plant Cell* 1, 141-150.

Ferl, R. J. & Nick, H. S. (1987) *J. Biol. Chem.* 262, 7947-7950.

Ferl, R. J. & Laughner, B. H. (1989) *Plant Mol. Biol.* 12, 357-366.

Ferl, R. J. (1990) *Plant Physiol.* 93, 1094-1101.

Fluhr, R., Moses, P., Morelli, G., Coruzzi, G. & Chua, N.-H. (1986) *EMBO J.* 5, 2063-2071.

Fox, P. C., Vasil, V., Vasil, I. K. & Gurley, W. B. (1992) *Plant Mol. Biol.* 20, 219-233.

Freedman, L. P., Luisi, B. F., Korzum, Z. R., Basavappa, R., Sibler, P. B. & Yamamoto, K. R. (1988) *Nature* 334, 543-546.

Fried, M. & Crothers, D. M. (1981) *Nucl. Acids Res.* 9, 6505-6525.

Fromental, C., Kanno, M. & Nomiya, H., & Chambon,P. (1988) *Cell* 54, 943-953.

Fromm, M. & Berg, P. (1983) *Mol. and Cell Biol.* 3, 991-999.

Fromm, H., Katagiri, F. & Chua, N.-H. (1989) *Plant Cell* 1, 977-984.

Galas, D. J. & Schmitz, A. (1978) *Nucl. Acids Res.* 5, 3157-3170.

Garner, M. M. & Revzin, A. (1981) *Nucl. Acids Res.* 9, 3047-3060.

Geuns, J. M. C. (1978) *Phytochemistry* 17, 1-14.

Geuns, J. M. C. (1983) *Biochem. Soc. Trans.* 11, 533-538.

Ghosh, P., Lebowitz, P., Frisque, R. J. & Gluzman, Y. (1981) *Proc. Natl. Acad. Sci. USA.* 78, 100-104.

Gillies, S., Morrison, S., Oi, V. & Tonegawa, S. (1983) *Cell* 33, 717-728.

Gilman, M. Z., Wilson, R. N. & Weinberg, R. A. (1986) *Mol. Cell. Biol.* 6, 4305-4316.

Goodwin, G. & Bustin, M. (1988) *Architecture of Eukaryotic Genes*, 187-205, VCH Publishers.

Grandbastien, M. A., Berry-Lowe, S., Shirley, B. W. & Meagher, R. B. (1986) *Plant Mol. Biol.* 7, 451-465.

Green, P. J., Kay, S. A. & Chua, N.-H. (1987) *EMBO J.* 6, 2543-2549.

Green, P. J., Yong, M.-H., Cuozzo, M., Kano-Murakami, Y., Silverstein, P. & Chua, N.-H. (1988) *EMBO J.* 7, 4036-4044.

Green, S., Kumar, V., Rhenlaz, I., Wahli, W. & Chambon, P. (1988) *EMBO J.* 7, 3037-3044.

Green, S. & Chambon, P. (1988) *Trends Genet.* 4, 309-314.

Grove, M. D., Spencer, G. F., Rohwedder, W. K., Mandava, N., Worley, J. F., Warthen, J. D., Steffens, G. L., Flippen-Anderson, J. L. & Cook, J. C., Jr. (1979) *Nature* 281, 216-217.

Guiliano, G., Pichersky, E., Malik, V.S., Timko, M.P., Scolnick, P.A., & Cashmore, A. R. (1988) *Proc. Natl. Acad. Sci. USA.* 85, 7089-7093.

Guiltinan, M. J., Marcotte, W. J. & Quatrano, R. S. (1990) *Science* 250, 267-271.

Hack, E. & Kemp, J. D. (1977) *Biochem. Biophys. Res. Commun. Science* 78, 785-791.

Hanas, J. S., Hazuda, D. J., Bogenhagen, D. F., Wu, F. Y.-H. & Wu, C.-W. (1983) *J. Biol. Chem.* 258, 14120-14125.

Harker, C. L., Ellis, T. H. N. & Coen, E. S. (1990) *Plant Cell* 2, 185-194.

Harpster, M. H., Townsend, J. A., Jones, J. D. G., Bedbrook, J. & Dunsmuir, P. (1988) *Mol Gen Genet.* 212, 182-190.

Harrison, M. J., Choudhary, A. D., Dubery, I., Lamb, C. J. and Dixon, Richard A. (1991) *Plant Mol Biol.* 16, 877-890.

Harrison, M. J., Lawton, M. A., Lamb, C. J., & Dixon, R. A. (1991) *Proc. Natl. Acad. Sci. USA.* 88, 2515-2519.

Heftmann, E. (1977) *Progr. Phytochem.* 4, 257-276.

Heidecker, G. & Messing, J. (1986) *Annu. Rev. Plant Physiol.* 37, 439-466.

Herrera-Estrella (1984) *Nature* 310, 115-120.

Hewitt, S., Hillman, J. R. & Knights, B. A. (1980) *New Phytol.* 85, 329-350.

Holsters, M., Hernalsteens, J. P., van Montagu, M. & Schell, J. (1982) *Molecular Biology of Plant Tumors*, 269-297. Academic Press.

Howard, E. A., Walker, J. C., Dennis, E. S. & Peacock, W. J. (1987) *Planta* 170,535-540.

Hu, J. C., O'Shea, E. K., Kim, P. S. & Sauer, R. T. (1990) *Science* 250, 1400-1403.

Jacobsen, K., Laursen, N. B., Jensen, K. O., Marcker, A., Poulsen, C. & Marcker, K. A.(1990) *Plant Cell* 2, 85-94.

Jensen, E. O., Marcker, K. A., Jeff, S. & de Bruijn, F. J. (1988) *EMBO.* 7, 1265-1277.

Jofuku, K. D., Okamuro, J. K. & Goldberg, R. B. (1987) *Nature* 328, 734-737.

Johnson, P. F. & McKnight, S. L. (1989) *Annu. Rev. Biochem.* 58, 799-839.

Jones, N. C., Rigby, P. W. & Ziff, E. B. (1988) *Genes and Dev.* 2, 267-281.

Joos, H., Caplan, A., Sormann, M., van Montagu, M. Schell, J (1983) *Cell* 32, 1057-1067.

Jordano, J., Almoguera, C. & Thomas, T. L. (1989) *Plant Cell* 1, 855-866.

Kadonaga, J. T. & Tjian, R. (1986) *Proc. Natl. Acad. Sci. USA.* 83, 5889-5893.

Kadonaga, J. T. (1991) *Methods in Enzymol.* 208, 10-45. Academic Press.

Karcher, S. J., DiRita, V. J. & Gelvin, S. B. (1984) *Mol Gen Gen.* 194, 159-165.

Karin, M., Haslinger, A., Holtgreve, H., Richards, R. I., Krauter, P., Westphal, H. & Beato, M. (1984) *Nature* 308, 513-519.

Katagiri, F., Lam, E. & Chua, N.-H. (1989) *Nature* 340, 727-730.

Kaulen, H., Schell, J. & Kreuzaler, F. (1986) *EMBO J.* 5, 1-8.

Khoury, G. & Gruss, P. (1983) *Cell* 33, 313-314.

Klee, H., Horsch, R. & Rogers, S. (1987) *Annu. Rev. Plant Physiol.* 38, 467-486.

Koes, R. E., Spelt, C. E., Mol, J. N. M. & Gerats, A. G. M. (1987) *Plant Mol Biol.* 10, 375-385.

Komro, C. T., DiRita, V. J., Gelvin, S. B. & Kemp, J. D. (1985) *Plant Mol. Biol.* 4, 253-263.

Koncz, C., De Greve, H., André, D., Deboeck, F., Van Montagu, M. & Schell, J. (1983) *EMBO J.* 2, 1597-1603.

Kononowicz, H., Wang, Y. E., Habeck, L., L & Gelvin, S., B. (1992) *Plant Cell* 4, 17-27.

Korber, H., Strizhov, N., Staiger, D., Feldwisch, J., Olsson, O., Sandberg, K. P., Schell, J. & Koncz, C. (1991) *EMBO J.* 10, 3983-3991.

Kuhlemeir, C., Fluhr, R., Green, P. J. & Chua, N.-H. (1987) *Genes and Dev.* 1, 247-255.

Laemmli, U. K. (1970) *Nature* 227, 680-685.

Lam, E., Benfey, P., Gilmartin, P. M., Fang, R.-X. & Chua, N.-H. (1989) *Proc. Natl. Acad. Sci. USA.* 86, 7809-7894.

Lam, E., Kano-Murakami, Y., Gilmartin, P., Niner, B. & Chua, N. (1990) *Plant Cell* 2, 857-866.

Lam, E., Katagiri, F. & Chua, N.-H. (1990) *J. Biol. Chem.* 265, 9909-9913.

Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988) *Science.* 240, 1759-1764.

Langridge, W. H. R., Fitzgerald, K. J., Koncz, C., Schell, J. & Szalay, A. A. (1989) *Proc. Natl. Acad. Sci. USA.* 86, 3219-3223.

Laudet, V., Hanni, C., Coll, J., Catzeflis, F. & Stehelin, D. (1992) *EMBO J.* 11, 1003-1013.

Lawton, M. A., Dean, S. M., Dron, M., Kooter, J. M., Kragh, K. M., Harrison, M. J., Yu, L., Tanguay, L., Dixon, R. A. & Lamb, C. J. (1991) *Plant Mol. Biol.* 16, 235-249.

Leisner, S. M. & Gelvin, S. B. (1988) *Proc. Natl. Acad. Sci. USA.* 85, 2553-2557.

Leisner, S. M. & Gelvin, S. B. (1989) *Plant Cell* 1, 925-936.

Leung, J., Fukuda, H., Wing, D., Schell, J. & Masterson, R. (1991) *Mol. Gen. Genet.* 230, 463-474.

Lin, S.-Y. & Riggs, A. D. (1974) *Proc. Nat. Acad. Sci. USA.* 71, 947-951.

Lois, R., Dietrich, A., Hahlbrock, K. & Schulz, W. (1989) *EMBO J.* 8, 1641-1648.

Mandava, N. B., Sasse, J. M. & Yopp, J. H. (1981) *Physiol. Plant* 53, 453-461.

Maniatis, T., Goodbourn, S. & Fischer, J. A. (1987) *Science* 230, 1237-1244.

Maxam, A. & Gilbert, W. (1980) *Methods in Enzymol.* 65, 499-560. Academic Press.

McKnight, S. L. & Kingsbury (1982) *Science* 217, 316-324.

McMorris, T. C. (1978) *Lipids* 13, 716-722.

Messens, E., Lenaerts, A., Van Montagu, M. & Hedges, R. W. (1985) *Mol. Gen. Genet.* 199, 344-348.

Minami, E. I., Ozeki, Y., Matsuoka, M., Koizuka, N. & Tanaka, Y. (1989) *Eur. J. Biochem.* 185, 19-25.

Mitchell, J. W., Mandava, N., Worley, J. F., Plimmer, J. R. & Smith, M. V. (1970) *Nature* 225, 1065-1066.

Mitra, A. & An, G. (1989) *Mol. Gen. Genet.* 215, 294-299.

Moore, D. D. (1990) *New Biol.* 2, 100-105.

Morelli, G., Nagy, F., Fraley, R. T., Rogers, S. G. & Chua, N.-H. (1985) *Nature* 315, 200-204.

Mosser, D. D., Kotzbauer, P. T., Sarge, K. D. & Morimoto, R. I. (1990) *Proc. Natl. Acad. Sci. USA.* 87, 3748-3752.

Nester, E. W. & Kosuge, T. (1981) *Ann. Rev. Microbiol.* 35, 531-565.

Nester, E. W., Gordon, M. P., Amasino, R. M. & Yanofsky, M. F. (1984) *Ann. Rev. Plant Physiol.* 35, 387-413.

O'Grady, K. (1993) Dissertation, Univ. of Florida.

Odell, J. T., Nagy, F. & Chua, N.-H. (1985) *Nature* 313, 810-812.

Odell, J. T., Knowlton, S., Lin, W. & Mauvais, C. J. (1988) *Plant Mol. Biol.* 10, 267-273.

Oeda, K., Salinas, J. & Chua, N.-H. (1991) *EMBO J.* 10, 1793-1802.

Ohl, S., Hedrick, S. A., Chory, J. & Lamb, C. J. (1990) *Plant Cell* 2, 837-848.

Okamuro, J. K. & Goldberg, R. B. (1989) *Biochemistry of Plants: A Comprehensive Treatise.* 15, 1-82. Academic Press.

Ondek, B., Shepard, A., & Herr, W. (1987) *EMBO J.* 6, 1017-1052.

Otten, L., De Greve, H., Hernalsteens, J. P., Van Montagu, M., Schieder, O., Straub, J. & Schell, J. (1981) *Mol. Gen. Genet.* 183, 209-213.

Ow, D. W., Jacobs, J. D. & Howell, S. H. (1987) *Proc. Natl. Acad. Sci. USA.* 84, 4807-4874.

Paonessa, G., Gounari, F., Frank, R. & Cortese, R. (1988) *EMBO J.* 7, 3115-3123.

Pelham, H. R. B. (1982) *Cell* 30, 517-528.

Perkins, N. D., Nicolas, R. H., Plumb, M. A. & Goodwin, G. H. (1989) *Nuc. Acids Res.* 17, 1299-1314.

Peterson, C. L. & Calame, K., L. (1987) *Mol. Cell Biol.* 7, 4194-4203.

Peterson, C., L. & Calame, K. (1989) *Mol. Cell. Biol.* 9, 776-786.

Power, R. F., Conneely, O. M. & O'Malley, B. W. (1992) *TiPS.* 13, 318-323.

Rasmussen, R., Benvegnu, D., O'Shea, E., Kim, P. S. & Albert, T. (1991) *Proc. Natl. Acad. Sci. USA.* 88, 561-564.

Rocha-Sosa, M., Sonnewald, U., Frommer, W., Stratmann, M., Schell, J., & Willmitzer, L. (1989) *EMBO J.* 8, 23-29.

Rubin, C. M. & Schmid, C. W. (1980) *Nucleic Acids Res.* 8, 4613-6419.

Ryder, T. B., Hedrick, S. A., Bell, J. N., Liang, X., Clouse, S. D. & Lamb, C. J. (1987) *Mol. Gen. Genet.* 210, 219-233.

Salomon, F., Deblaere, R., Leemans, J., Hernalsteens, J.-P., van Montagu, M. & Schell, J. (1984) *EMBO J.* 3, 141-146.

Sambrook, J., Fritsch, Edward F., and Maniatis, Thomas (1989) Molecular Cloning A laboratory manual. 2nd. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Sanger, M., Daubert, S. & Goodman, R. M. (1990) *Plant Mol Biol*. 14, 433-443.

Sassone-Corsi, P. & Borrelli, E. (1986) *Trends in Genet*. 215-219.

Sawadago, M. & Roeder, R. G. (1985) *Cell* 43, 165-175.

Schindler, U., Terzaghi, W., Beckmann, H., Kadesch, T. & Cashmore, A. (1992) *EMBO J*. 11, 1257-1289.

Schindler, U., Menkens, A. E., Beckman, H., Ecker, J. & Cashmore, A. R. (1992) *EMBO J*. 11, 1261-1273.

Schirm, S., Jiricny, J., & Schaffner, W. (1987) *Genes and Devel*. 1, 65-74.

Schmidt, R. J., Burr, F. A., Aukerman, M. J. & Burr, B. (1990) *Proc. Natl. Acad. Sci. USA*. 87, 46-50.

Schoffl, F. & Key, J. L. (1982) *J. Mol Appl Genet*. 1, 301-306.

Schroder, J., Schroder, G., Huisman, H., Schilperoort, R. A. & Schell, J. (1981) *FEB Lett*. 129, 166-168.

Schulze-Lefert, P., Dangl, J. L., Becker-Andre, M., Hahlbrock, K. & Schulz, W. (1989) *EMBO J*. 8, 651-656.

Schulze-Lefert, P., Becker-Andre, M., Schulz, W., Hahlbrock, K. & Dangl, J. L. (1989) *Plant Cell* 1, 707-714.

Schwartz, D. (1966) *Proc. Natl. Acad. Sci. USA*. 56, 1431-1436.

Serfling, E., Jasin, M. & Schaffner, W. (1985) *Trends in Genet*. 224-230.

Shaw, C. H., Carter, G. H., Watson, M. D. & Shaw, C. H. (1984) *Nucl. Acids Res.* 12, 7831-7846.

Singh, H., Sen, R., Baltimore,D., & Sharp, P.A. (1986) *Nature* 319, 154-158.

Singh, K., Tokuhisa, J. G., Dennis, E. S. & Peacock, W. J. (1989) *Proc. Natl. Acad. Sci. USA.* 86, 3733-3737.

Singh, K., Dennis, E., Ellis, J., Llewellyn, D. J., Tokuhisa, J. G., Wahleithner, J. A. & Peacock, W. J. (1990) *Plant Cell* 2, 891-903.

Slama, K. (1980) *Biochem. Physiol. Pflanzen.* 175, 177-193.

Spiker, S. (1984) *J. Biol. Chem.* 259, 12007-12013.

Staiger, D., Kaulen, H. & Schell, J. (1989) *Proc. Natl. Acad. Sci. USA.* 86, 6930-6934.

Staiger, D., Becker, F., Koncz, C. & Palme, K. (1991) *Eur. J. Biochem.* 199, 519-527.

Strabala, T., Crowell, D. N. & Amasino, R. M. (1993) *Plant Mol. Biol.* 21, 1011-1021.

Struhl, K. (1989) *Trends Biochem. Sci.* 14, 137-140.

Tabata, T., Takase, H., Takayama, S., Mikami, K., Nakatsuka, A., Kawata, T., Nakayama, T. & Iwabuchi, M. (1989) *Science* 245, 965-967.

Tabata, T., Nakayama, T., Mikami, K. & Iwabuchi, M. (1991) *EMBO J.* 10, 1459-1467.

Teeri, T. H., Lehvaslaiho, H., Frank, M., Uotila, J., Heino, P., Palva, E. T., van Montagu, M. & Herrera-Estrella, L. (1989) *EMBO J.* 8, 343-350.

Thomashow, M., Nutter, R., Montoya, A. L., Gordon, M. P. & Nester, E. (1980) *Cell* 19, 729-739.

Thornburg, R. W., An, G., Cleveland, T. E., Johnson, R. & Ryan, C. A. (1987) *Proc. Natl. Acad. Sci.* USA. 84, 744-748.

Timko, M. P., Kausch, A. P., Castresana, C., Fassler, J., Herrera-Estrella, L., Van den Broeck, G., Van Montagu, M., Schell, J. & Cashmore, A. R. (1985) *Nature* 318, 579-582.

Tokuhisa, J. G., Singh, K., Dennis, E. S. & Peacock, W. J. (1990) *Plant Cell* 2, 215-224.

Truss, M., Chalepakis, G. & Beato, M. (1990) *Proc. Natl. Acad. Sci. USA.* 87, 7180-7184.

Tsai, S. Y., Sagami, I., Wang, H., Tsai, M.-J. & O'Malley, B. W. (1987) *Cell* 50, 701-709.

Velten, J., Velten, L., Hain, R. & Schell, J. (1984) *EMBO J.* 3, 2723-2730.

Vinson, C. R., Sigler, P. B. & McKnight, S. L. (1989) *Science* 246, 911-916.

Vodkin, L. O., Rhodes, P. R. & Goldberg, R. B. (1983) *Cell* 34, 1023-1031.

Walker, J. C., Howard, E. A., Dennis, E. S. & Peacock, W. J. (1987) *Proc. Natl. Acad. Sci. USA.* 84, 6624-6628.

Wang, L.-H., Tsai, S. Y., Sagami, I., Tsai, M.-J. & O'Malley, B. W. (1987) *J. Biol. Chem.* 262, 16080-16086.

Wang, L.-H., Tsai, S. Y., Cook, R. G., Beattie, W. G., Tsai, M.-J. & O'Malley, B. W. (1989) *Nature* 340, 163-168.

Weiher, H., Konig, M. & Gruss, P. (1983) *Science* 219, 626-631.

Weising, K., Schell, J. & Kahl, G. (1988) *Ann. Rev. Genet.* 22, 421-477.

Weisshaar, B., Armstrong, G. A., Block, A., da Costa e Silva, O. & Hahlbrock, K. (1991) *EMBO J.* 10, 1777-1786.

White, F. F. & Nester, E. W. (1980) *J. Bacteriol.* 141, 1134-1141.

Williams, M. E., Foster, R. & Chua, N.-H. (1992) *Plant Cell* 4, 485-496.

Willmitzer, L., W., S. & Schell, J. (1981) *Nucl. Acids Res.* 9, 4801-4812.

Willmitzer, L., Simons, G. & Schell, J. (1982) *EMBO J.* 1, 139-146.

Willmitzer, L., Sanchez-Serrano, J., Buschfeld, E. & Schell, J. (1982) *Mol. Gen. Genet.* 186, 16-22,

Winter, J. A., Wright, R. L. & Gurley, W. B. (1984) *Nucl. Acids Res.* 12, 2391-2406.

Wu, C., Wilson, S., Walker, B., Dawid, I., Paisley, T., Zimarino, V. & Ueda, H. (1987) *Science* 238, 1247-1252.

Xiao, H. & Lis, J. T. (1988) *Science* 239, 1139-1142.

Yopp, J. H., Colclasure, G. C. & Mandava, N. (1979) *Physiol. Plant.* 46, 247-254.

Yopp, J. H., Mandava, N. B. & Sasse, J. M. (1981) *Physiol. Plant.* 53, 445-452.

Zinn, K. & Maniatis, T. (1986) *Cell* 45, 611-618.

BIOGRAPHICAL SKETCH

I was born in Southhampton county Virginia in September, 1942. My educational background includes the completion of primary and secondary schooling in Norfolk, Virginia. The B.S. Degree in Biology was granted by Norfolk State University, Norfolk, Virginia in June 1964. The M.S. Degree in Reading Education was granted by Syracuse University, Syracuse, New York in 1974. Most of my adult life has been spent teaching high school biology while brief periods were spent in hospital employ as a Registered Medical Technologist and teaching reading and study skills from the primary grades through the college level.

Prior to my enrollment at the University of Florida, I was teaching high school biology in Elkhart, Indiana. The field of molecular biology was becoming increasingly more advanced and intriguing to me. At this time I became aware of the McKnight Black Doctoral Fellowship. Receipt of this fellowship would allow me to learn more about molecular biology and to fulfill a long held dream to obtain a Doctor of Philosophy Degree in order to teach and to conduct research at the college level. I applied, and out of a pool of thousands, I was awarded the fellowship.

Entering the University of Florida and succeeding was a major challenge for me. My first year was spent taking undergraduate courses which allowed me to update my knowledge in the field and also served as a barometer measuring my academic capabilities at that time in my life. Upon the successful completion of the first year, I embarked upon graduate course work as outlined by the department of Microbiology and Cell Science and joined the Gurley laboratory. Since then, I have successfully completed my course work, passed the written and oral comprehensive examinations, and have become a candidate for the Doctor of Philosophy Degree. I am presently a member of the American Association for the Advancement of Science, the National Society of Plant Physiology and the American Association of University Women Professors. As a graduate student, I have made three presentations at the Annual University of Florida Plant Molecular and Cellular Biology Workshops held at the University of Florida Marine Research Center in St. Augustine, Florida. I also presented a poster abstract at the International Plant Molecular Biology conference held in Tucson Arizona, 1990.

I have maintained a happy marriage to Howard Adams, Executive Director of the GEM program based at the University of Notre Dame and have reared one female child, Stephanie, who earned a Masters degree in Systems Engineering and is presently employed at North Carolina State University as the Director of the Minority Engineering Program.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

William B. Gurley,
Chairman
Associate Professor of
Microbiology and Cell
Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Francis C. Davis
Associate Professor of
Microbiology and Cell
Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Robert J. Ferl
Professor of
Horticultural Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Donald R. McCarty
Associate Professor of
Horticultural Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Robert R. Schmidt
Graduate Research
Professor of Microbiology
and Cell Science

This dissertation was submitted to the Graduate Faculty in the College of Agriculture and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August, 1993

Dean, College of
Agriculture

Dean, Graduate School